Investigation of virulence gene regulation in *Streptococcus pneumoniae*

Thesis

INVESTIGATION OF VIRULENCE GENE REGULATION IN *STREPTOCOCCUS PNEUMONIAE*

Carolyn Hemsley

MA BM BCh MRCP DipPath

2004

Submitted to fulfil the requirements for the degree of PhD, Life Sciences, Open University. Weatherall Institute of Molecular Medicine, Oxford (sponsoring establishment) and Tufts university, Boston, MA, USA (collaborating establishment).

ABSTRACT

Streptococcus pneumoniae is an important human pathogen in all age groups worldwide that causes a variety of diseases, ranging from life threatening septicaemia and meningitis to less severe sinusitis and otitis media. The factors that determine the virulence of *S. pneumoniae* are very complex but a key aspect of the organism's disease causing potential is the ability of the bacteria to regulate virulence factor expression and activity. In this study two main approaches were taken to investigate virulence gene expression in *S. pneumoniae*. Firstly, the feasibility of Recombinase based *In vivo* Expression Technology, RIVET, for use in *S. pneumoniae* to study gene expression *in vitro*, and then *in vivo* was assessed. However, the system was found to be unsuitable for use in this study. Secondly, the requirement for and the role of virulence gene regulators identified by Signature Tagged Mutagenesis were investigated. The requirement for different virulence gene regulators varied according to the murine model of infection used. Two of the regulators, MgrA and RlrA, were essential for nasopharyngeal carriage and production of pneumonia in mice by serotype 4 *S. pneumoniae*. Both were shown to control the transcription of genes of a newly described pathogenicity islet, PPI2, encoding RlrA and proteins predicted to act at the bacterial cell surface. The PPI2 genes *rlrA* and *rrgA* were shown to be required for adhesion of serotype 4 *S. pneumoniae* to human epithelial cells and PPI2 gene expression was affected by the gaseous composition of the growth environment in an MgrA dependent manner. The distribution of MgrA, RlrA and PPI2 varied between clinical *S. pneumoniae* isolates emphasizing the likelihood of a different repertoire of virulence genes and regulators amongst different serotypes and strains of this important human pathogen.

ACKNOWLEDGEMENTS

I would like to thank my supervisors and mentors Andrew Camilli, Derrick Crook, Richard Moxon and David Holden, without whose teaching, guidance and discussions I could never have completed this work. I would like to thank members of both laboratories in which I worked and in particular Susan Butler, who has been a good friend and provided many suggestions and practical advice towards the project. I also thank my friends and family for encouragement. My greatest thanks are reserved for my husband: for his never-ending support through my research fellowship. He kept me calm and focused through this period of many changes in our lives.

This thesis is dedicated to Thomas Frederick Archie Hettiaratchy and Alice Emmeline Daisy Hettiaratchy. I produced three 'babies' in the three years of my PhD. The thesis was one and they were the others.

LIST OF ABBREVIATIONS

AOM	acute otitis media
CBP	choline binding protein
CFU	colony forming unit
ChoP	phosphorycholine
CI	competition index
CM	cholramphenicol
COI	cancelled out index
CSP	competence stimulating peptide
ECM	extracellular matrix
GAS	group A streptococcus
GFP	green fluorescent protein
HK	histidine kinase
IS	insertion sequence
MLST	multi-locus sequence typing
MSCRAMM	microbial surface components recognizing adhesive matrix molecules
ORF	open reading frame
PAI	pathogenicity island
PBP	penicillin binding protein
PCR	polymerase chain reaction
pIgR	polymeric immunoglobulin receptor
PPI1/2	pneumococcal pathogenicity island 1/2
PRSP	penicillin-resistant *S. pneumoniae*
RBS	ribosome binding site
RIVET	recombinase-based *in vivo* expression technology

RPA	ribonuclease protection assay
rPAF	receptor for platelet activating factor
RR	response regulator
SM	streptomycin
Spc	Spectinomycin
ST	sequence type
STM	Signature Tagged Mutagenesis
TCSTS	two component signal transduction system
THY	Todd Hewitt broth plus 5% yeast extract

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER 1 INTRODUCTION

CHAPTER 2 MATERIALS AND METHODS

CHAPTER 3 RIVET IN *S. PNEUMONIAE*

CHAPTER 6 FUNCTION AND DISTRIBUTION OF *RLRA* PATHOGENICITY ISLET

CHAPTER 7 DISCUSSION

CHAPTER 1 INTRODUCTION

1.1 *STREPTOCOCCUS PNEUMONIAE*

Streptococcus pneumoniae is a Gram-positive coccus that characteristically grows in pairs or chains in liquid medium. It was originally isolated in 1880 by Louis Pasteur and G. M. Sternberg though was not assigned a name, *Diplococcus pneumoniae,* until the 1920s. This name (probably due to its characteristic pairing seen in sputum of patients with pneumonia) was changed to *S. pneumoniae* in 1974 (Watson *et al.*, 1993; Winslow *et al.*, 1920). A brief review of the history of *S. pneumoniae* highlights the interrelationship between the study of this organism and the developments in microbiology and infectious disease including the discovery of transformation, antibiotic use and vaccine development to name a few. Yet after more than a century since its initial isolation *S. pneumoniae* remains a major cause of morbidity and mortality worldwide.

1.1.1 Carriage, disease and epidemiology

Although an important human pathogen, *S. pneumoniae* usually resides as a component of the normal microflora of the nasopharynx. This asymptomatic colonization occurs shortly after birth (Austrian, 1986) and continues intermittently throughout life. It is a highly dynamic process; acquisition is followed by carriage for a period of weeks or months and then loss and recolonisation with other strains (Gray *et al.*, 1980). Younger children carry single strains for months whereas older toddlers and adults have a more rapid flora turnover in matter of weeks. At any one time and with appropriate culture methods, the organism can be detected in 20-30% of healthy adults and up to 40% of healthy children (Ghaffar *et al.*, 1999; Leino *et al.*, 2001). Actual pneumococcal disease

is considered to occur subsequent to nasopharyngeal carriage (Gray *et al.*, 1980; Obaro and Aaberge, 2002) but is a relatively rare event with only a small number of those colonized proceeding to symptomatic disease. This can vary in severity from mild otitis media and sinusitis, to pneumonia and lethal septicaemia and meningitis. For all forms of pneumococcal disease those at highest risk are children less than 5 years old, the elderly and people with chronic underlying medical conditions, including HIV. *S. pneumoniae* is the commonest cause of community-acquired bacterial pneumonia in adults and children, with an incidence ranging from 1.7 per 1000 person-years to as great as 19.0 in those with risk factors (Mangtani *et al.*, 2003; Simberkoff *et al.*, 1986). There is a significant mortality associated with invasive disease, highest in developing countries but still of major importance in developed countries. Studies over the last 10 years vary in absolute figures from 10-20 cases of invasive disease per 100,000 individuals (Sleeman *et al.*, 2001) across Europe and USA. This difference most likely reflects differences in reporting and clinical practice between countries. However, even the milder disease manifestations place a large burden on medical services. *S. pneumoniae* has been reported to account for 30-50% of otitis media episodes in the USA, resulting in 7 million otitis media-related visits per year to paediatricians (1997). One US-based study determined that 62% of children have an episode of acute otitis media by one year of age, with most cases occurring between the ages of 6 to 12 months (Teele *et al.*, 1989).

1.1.2 *S. pneumoniae* serotypes and vaccines

It is clear that not all strains of *S. pneumoniae* are created equal with respect to their ability to cause invasive disease. There are 90 serotypes of *S. pneumoniae* defined by variations in the immunochemistry of their polysaccharide capsule. Geographical, age-

related differences and differences in disease manifestations with certain serogroups have led to the suggestion that each strain be considered as a separate pathogen from an epidemiological standpoint (Hausdorff *et al.*, 2000a; Hausdorff *et al.*, 2000b).

The polysaccharide capsule is arguably the organism's most important virulence factor. Isolates lacking a capsule rarely cause invasive disease in humans and are greatly attenuated in virulence in animal models of infection (Malley *et al.*, 2001). Antibodies against capsule are highly protective against invasive disease. For these reasons initial pneumococcal vaccine efforts were directed against capsular polysaccharide and the capsule still remains the target of anti-pneumococcal vaccines in present day use. The capsule varies in chemical composition between serotypes (Kamerling, 2000) and its actual composition and structure is considered important in the disease-causing potential of the organism. There is an apparent serotype-specific difference in invasive potential. A recent epidemiological study compared 501 mixed carriage and invasive strains of *S. pneumoniae*. The carriage isolates had been obtained from nasopharyngeal aspirates from healthy young children and the invasive strains were clinical isolates recovered from normally sterile sites in children with proven pneumococcal disease. Serotypes 4 and 1 were found to be overrepresented among isolates from invasive disease. At the other extreme one serotype 3 clone appeared to have a low invasive potential being almost invariably recovered from carriage (Brueggemann *et al.*, 2003). In fact more than 80% of serious illness is caused by only 15 serotypes (Hausdorff *et al.*, 2000a; Hausdorff *et al.*, 2000b; Scott *et al.*, 1996). Some serotypes are particularly associated with disease in particular populations such as children or are associated with habouring antibiotic resistance (McGee *et al.*, 2001; Scott *et al.*, 1996).

The rank order of serotypes recovered from invasive disease in developed countries has been used as a guide to the development of both the 23-valent polysaccharide vaccine and the newer conjugate vaccines. An original 14-valent polysaccharide was licensed in 1977 covering the 14 serotypes causing 75% of invasive disease (Austrian, 1977). This was extended to the 23-valent version in 1983 which encompasses the 23 pneumococcal serotypes that account for 90% of isolates recovered from sterile sites (1997). This vaccine is widely used today and is recommended for use in people over the age of 60 and in high-risk individuals aged more than 5 years. However the polysaccharide antigens are poorly immunogenic in children less than 2 years of age, where a large burden of disease lies. Following on from the success of the conjugate HIB vaccine in this age group protein-polysaccharide conjugate pneumococcal vaccines have been developed for use in children less than 5 years of age. A 7-valent vaccine, licensed for use in USA in 2000, now constitutes part of the routine vaccination schedule in that country. The conjugate vaccine in use, Prevnar, covers the 7 serotypes; 4, 6B, 9V, 14, 18C, 19F and 23F, which constitute 75% of those causing invasive disease in either USA or UK and Europe but only 45-60% of invasive serotypes in Latin America, Africa and parts of Asia (Hausdorff *et al.*, 2000b; Sleeman *et al.*, 2001). Inclusion of a further 4 serotypes in an 11-valent vaccine increases coverage to greater than 75% in all areas worldwide but this has serious cost implications (Hausdorff *et al.*, 2000b). The 7-valent vaccine is already considered too expensive by most developing countries to enable mass vaccination.

Although it is still early since the introduction of the newer pneumococcal conjugate vaccines studies have already shown that they provide some protection against *S. pneumoniae* disease in the targeted age group in certain populations. Two studies from

the USA reported a clear benefit in bacteraemic disease with a 93.9% (95% CI 79.6-98.5) and 78.6% (95% CI 9.4-95.1) efficacy in preventing bacteraemia (Black *et al.*, 2000; O'Brien *et al.*, 2003). The results for preventing pneumonia and otitis media are less impressive (Klugman, 2001a). Finnish trials showed only a 56% (95% CI 44-66) or 57% (95% CI 44-67) efficacy for prevention of acute otitis media (AOM) (Eskola *et al.*, 2001; Kilpi *et al.*, 2003). However it must be recognized that even a small reduction in incidence of disease can equate to a huge impact in terms reduction of actual numbers of cases. There are equally interesting data coming from studies examining the efficacy of these vaccines and their effect on non-vaccine serotype strains and nasopharyngeal carriage. There is a reduction in carriage of vaccine serotype strains but also seemingly an increase in carriage of non-vaccine serotypes (Klugman, 2001b; Mbelle *et al.*, 1999; Obaro *et al.*, 1996). This has long-term implications for the efficacy of the vaccines. In one Finnish study the increase in carriage was also reflected in an increase of 33% in AOM by non-vaccine serotypes (Eskola *et al.*, 2001). Replacement of carried pneumococci by non-vaccine serotype strains could occur by many different mechanisms. The most likely explanation is that children who are immunized are at increased risk of colonization with non-vaccine serotypes. If they are exposed to both non-vaccine and vaccine serotypes but are immune to the acquisition of latter, secondary to immunization, they will be preferentially colonized by the former. Alternatively it may be that replacement colonization represents the unmasking of sub-dominant serotypes, previously carried in low numbers. A final possibility is that the host-immune response selects for clones that have undergone *in vivo* recombination of the capsule genes and thus switched capsule serotype. As discussed later, this capsular switching does occur *in vivo* but is considered to be a very rare event. Whatever the mechanism there is evidence that the widespread introduction of conjugate vaccines

will, with time, affect carriage serotype distribution and this may have a knock-on effect in those serotypes causing invasive disease.

1.1.3 Upsurgence of antibiotic resistance

Hand in hand with the apparent success and concerns over longterm failure of new pneumococcal vaccines in disease prevention, disease treatment is also providing new challenges. Penicillin has historically been the antibiotic of choice for treatment of pneumococcal disease. Yet as early as 1941, only 12 years after the discovery of penicillin, resistance to the drug could be detected in clinical isolates (Abraham *et al.*, 1941) and since the 1960s there has been a dramatic worldwide increase in resistance to penicillin and other antibiotics, including macrolides, chloramphenicol and tetracyclines in clinical isolates (Doern *et al.*, 2001; Huebner *et al.*, 2000; Whitney *et al.*, 2000). The mechanisms by which strains become β-lactam resistant have been well studied. Penicillin and other β-lactam drugs bind to and inhibit penicillin-binding proteins (PBPs), which normally catalyse important steps in the assembly of peptidoglycan in the bacterial cell wall. Resistance is due to alterations in these PBPs, leading to a decreased binding affinity both for drug and normal substrate (Hakenbeck *et al.*, 1999). *S. pneumoniae* contains six PBPs. All can occur as low affinity variants (Hakenbeck *et al.*, 2000). Interestingly these low affinity forms of PBPs are recognized as the products of DNA transformation and recombination events. They are encoded by mosaic genes that contain sequence blocks highly divergent from those of sensitive strains and as such appear to be a result of gene transfer not only among pneumococcal strains but also from other streptococcal species. Low affinity forms of PBP 2b and 2x confer low-level resistance and are prerequisites for high-level β-lactam resistance.

Penicillin-resistant strains compensate for the changes in affinity for the cell wall building blocks by activation of genes *murM* and *murN* which produce branched chain cell wall peptides rather than the typical linear structures. The capacity to produce these branched chain cell wall precursors is a prerequisite for expression of PBPs with lowered affinity for penicillin as interruption of *murMN* leads to complete inhibition of expression of penicillin resistance (Filipe *et al.*, 2001, 2002). The *murMN* genes or their products could be new drug targets for penicillin-resistant *S. pneumoniae* (PRSP).

It was realized in the 1990s that PRSP were more likely to be resistant to other antibiotic groups such as macrolides, teracyclines, chloramphenicol and most recently quinolones. As currently two thirds of clinical isolates in the USA have intermediate or high level resistance to penicillin, 40-60% of these are also resistant to at least one other antibiotic and this mirrors the trend worldwide. Pneumococcal disease management therefore remains an ever increasing challenge (Garau, 2002).

1.2 EVOLUTION OF *S. PNEUMONIAE*

1.2.1 Transformation and competence and in *S. pneumoniae*

Experiments performed by Griffith in the 1920s showed that intraperitoneal injection of live unencapsulated serotype 2 *S. pneumoniae* mutants together with heat-killed serotype 1 encapsulated *S. pneumoniae* led to the emergence of viable type1 encapsulated bacteria resulting in death of the mice (Griffith, 1928). Griffith termed this process 'transformation'. These initial findings were not fully explained until the 1940s when Avery and colleagues demonstrated that deoxyribonucleic acid (DNA) was the transforming agent and contained the genetic material encoding the capsular phenotype (Avery *et al.*, 1944). With the use of antibiotic resistance markers it was possible to

demonstrate the transfer of genetic material from one bacterium to another by freshly released DNA (Hotchkiss, 1951) and later demonstrate that such cell to cell transformations occurred both in infected animals and between bacteria grown in lab medium (Ottolenghi and Hotchkiss, 1962; Ottolenghi and MacLeod, 1963; Ottolenghi-Nightingale, 1969).

It is now appreciated that this uptake of exogenous native DNA and subsequent homology-dependent integration into the recipient chromosome (transformation) is one of the simplest modes of exchange of genetic material. *S. pneumoniae* is naturally transformable. During the transformation process itself exogenous double-stranded DNA binds to the cell membrane and is subjected to single-stranded nicks (Lacks and Greenberg, 1976). One strand is then transported across the cell membrane and if sufficiently similar in nucleotide sequence to a chromosomal counterpart is incorporated into the recipient's genome by homologous recombination (Fox and Allen, 1964; Prudhomme *et al.*, 2002). Competence, a physiological state in which exogenous DNA can be internalized, is a prerequisite for transformation in *S. pneumoniae*. It arises in exponentially growing cultures at a critical cell density and is dependent on a quorum sensing system. The cell-to-cell signal is called competence stimulating peptide, CSP. It is a small peptide processed from a precursor (ComC) (Havarstein *et al.*, 1995), and secreted by the ABC transporter ComAB (Hui *et al.*, 1995). CSP is sensed by a receptor histidine protein kinase, located in the cell membrane (ComD), which is linked to a response regulator (ComE). Binding of CSP activates a phosphorelay resulting in the phosphorylation of response regulator ComE. This phosphorylated form of ComE is a transcriptional activator and induces the transcription of numerous genes. It binds

specifically to a target site consisting of two 9 bp imperfect direct repeats separated by a stretch of 12 nucleotides in the promoter region of its target genes (Ween *et al.*, 1999). These include the *comCDE* operon and a large number of genes, encoding factors with functions in binding, uptake, processing and integration of exogenous DNA (Campbell *et al.*, 1998; Mortier-Barriere *et al.*, 1998; Pestova and Morrison, 1998). Regulation also depends on a novel sigma factor (ComX) itself controlled by the autostimulatory CSP sensing system (Morrison and Lee, 2000). Initial microarray studies had shown that the whole competence regulon is made of two classes of genes, early (e.g. *comCDE*) and late (Peterson *et al.*, 2000; Rimini *et al.*, 2000). The early genes rely on ComE for expression whereas the late genes depend on ComX (Lee and Morrison, 1999; Luo and Morrison, 2003). More recent comprehensive microarray studies have expanded the identification of competence induced genes to over 100 genes including a further pool of genes with a delayed expression profile and a group of genes that are repressed during the CSP response (Dagkessamanskaia *et al.*, 2004; Peterson *et al.*, 2004).

1.2.2 Transformation and Recombination and implications for evolution and population structure in *S. pneumoniae*

It is clear that the process of transformation allows the exchange of genetic material within a single species and potentially between species. The cellular machinery necessary for transformation has been found to be present in most *S. pneumoniae* isolates (Ramirez *et al.*, 1997). Recent molecular typing tools such as multilocus sequence typing (MLST) and statistical methods have allowed a quantitative estimation of the impact of recombination versus mutation on population structure (Enright and Spratt, 1998; Feil *et al.*, 2000). For *S. pneumoniae* it is estimated that individual

housekeeping genes are ~10 times more likely and individual nucleotides are ~50 times more likely to evolve by recombination than mutation and hence it has been suggested that specific clones of *S. pneumoniae* evolved predominantly by recombination (Feil *et al.*, 2000). This has been questioned specifically for serotype 6 strains (Ashley Robinson *et al.*, 2002) where the authors report that both recombination and mutation have played important roles in the evolution of serotype 6A clones whereas recombination alone was the dominant mechanism for evolution of serotype 6B. This study used MLST to identify major clones amongst 212 carriage and invasive strains. Serotype 6A isolates were more frequently isolated from the carriage state and 6B isolates from invasive disease. The original MLST work in *S. pneumoniae* referred to above was performed on 274 invasive strains only. It is possible that evolutionary pressures on *S. pneumoniae* differ between carriage and invasive strains resulting in bias towards different systems predominating in each.

A single pneumococcal serotype typically comprises a number of genetically divergent clones as assessed by MLST (Brueggemann *et al.*, 2003; Takala *et al.*, 1996) (Enright and Spratt, 1998). At least some of this may be as a consequence of the horizontal transfer of the capsular biosynthetic genes into new lineages. In all but one serotype (serotype 37) the capsule biosynthetic loci are located in the same position of the *S. pneumoniae* chromosome, between *dexB* and *aliA* genes and are organized into cassettes (Garcia *et al.*, 2000). Most capsule cassettes also share a common arrangement. The first four genes (*capABCD)* share some degree of similarity and the remaining genes, predicted to be involved in capsule biosynthesis are serotype-specific. The regions flanking the cassette are also similar between serotypes. This arrangement is ideal for facilitating the transfer of the entire capsule locus between pneumococcal

strains resulting in serotype switching, which has serious implications for capsule-based vaccines. Evidence that *S. pneumoniae* strains can indeed change their serotype *in vivo* was first obtained from the analysis of populations of antibiotic resistant strains. Isolates were identified that differed in serotype but that were indistinguishable by multilocus enzyme electrophoresis, and restriction endonuclease cleavage electrophoretic profiling of penicillin-binding protein (PBP) genes, demonstrating them to be members of a single clone (Coffey *et al.*, 1991). Subsequent reports of serotype variants of antibiotic resistant clones including molecular studies have shown that indeed serotype changes occur by recombinational events replacing the capsular cassette of a recipient with that of the donor (Coffey *et al.*, 1998; Coffey *et al.*, 1999). A recent carriage study sampling children in the Oxford region showed that 68 out of 100 children carried multiple serotypes during the first two years of life (Meats *et al.*, 2003). In all cases isolates from the same children that varied in serotype also varied in clonal type indicating the acquisition of a different pneumococcal strain rather than a change specifically in capsular type secondary to a recombinational event. This indicates that serotype replacement by the afore mentioned type of recombinational event is not so common that it can be detected in cohorts over relatively short periods of time.

1.2.3 Pathogenicity Islands and *S. pneumoniae*

Microarray technology has allowed the investigation of variation within the entire chromosome between strains of *S. pneumoniae*. One study comparing 20 different *S. pneumoniae* isolates (representing 10 different serotypes) by hybridization of chromosomal DNA to an oligonucleotide array representing the serotype 4 genome showed that only 75% of the loci were conserved in all isolates (Hakenbeck *et al.*, 2001). A total of 470 genes showed variation in at least one of the strains and between

8-11% of genes varied in each strain. It appears therefore that the genes contained by one particular strain represent only a limited snapshot of the genetic information contained within the pneumococcal species as a whole. There is a stable pool of DNA, which is common to all strains, and a separate pool of variable DNA, which contains amongst other things the capsule biosynthetic loci.

With the availability of genome sequences of numerous bacterial species it has become apparent that many genomes take this form and consist of a core sequence, which encodes housekeeping functions, and additional variable sequences. Yet even before the advent of genome sequence availability specific sequences had been discovered in numerous species that differ from the rest of the genome in their G+C content and codon usage and that appear to have been acquired by horizontal transfer. These so called islands encode diverse functions ranging from biosynthetic pathways, antibiotic resistance and properties involved in microbial fitness. A subset of genomic islands contain genes involved in bacterial virulence and are called pathogenicity islands (PAIs). The first PAI was described in uropathogenic *E. coli* (Hacker *et al.*, 1983) but many species have DNA segments that share typical features of a PAI (Hacker *et al.*, 1997; Hacker and Kaper, 2000). A PAI is defined as a region of the genome, which is absent or sporadically distributed in non-pathogenic close relatives, that carries one or more virulence factor. The overall nucleotide composition of the island often differs from the core genome in G+C content and codon usage, reflecting the generation of PAIs by horizontal gene transfer. This G+C difference may not be observed if the DNAs of the donors and the recipients have a similar G+C content. PAIs are often flanked by sequences defining the edges such as direct repeats that could arise from integration of a transposon or bacteriophage, but often they do not represent

homogenous pieces of DNA but rather are made up of mosaic-like structures which have been generated by a multi-step process and thus boundaries can become obscured and harder to delineate with time.

S. pneumoniae appears to possess PAIs though only two islands have formally been described (Brown *et al.*, 2004; Hava *et al.*, 2003a). Both are present in serotype 4 (TIGR4 strain). Amongst the variable genes identified by comparative genomics were 10 gene clusters where the hybridization signal was uniformly low, encompassing 130 loci. Seven of these contain one or more genes encoding transposases and one resembles a prophage suggestive of mobile elements (Hakenbeck *et al.*, 2001). The sequences and exact positions of these clusters were not reported in this study. A second characterization of inter-strain genomic difference by microarray identified 9 TIGR4 gene clusters that did not hybridize with D39 or R6 genomic DNA, 6 of these had atypical nucleotide composition. Not surprisingly the capsule locus was common to both studies, so was an area with an IgA protease paralog (SP0071) and a gene encoding a choline binding protein (SP0061) (Tettelin *et al.*, 2001).

Which components of the *S. pneumoniae* genome are important for determining virulence? Are these components common to all strains or do they show inter-strain variability? How many virulence determinants are encoded for by PAIs? These questions are hard to answer with our present knowledge. Historically researchers have focused on capsule type to define disease-inducing potential and it is true that there is an apparent serotype-specific difference in invasive potential, but serotype is by no means the only contributing factor. Experiments examining the colonizing and disease forming potential of serotype 2, 3 and chimeric pneumococcal mutants with capsular

switching demonstrated that this was the case. For example, wild-type serotype 3 strain and a capsule switched mutant (genetically serotype 2 in all but capsule locus) behaved similarly with respect to nasopharyngeal colonization but the wild-type serotype 3 strain survived less well than the capsule switched mutant in the lung (Kadioglu *et al.*, 2002) showing that there must be other genomic differences outside the capsular locus contributing to survival. What we can say from these and other experiments is that a combination of capsule and genetic background determines virulence and these virulence-determining factors do not appear to be common to all strains. Factors that are required for virulence in one strain may not be needed in another and presently only a few of the known virulence factors have actually been shown to be on PAIs.

1.3 VIRULENCE FACTORS IN *S. PNEUMONIAE*

Analysis of virulence factors used by pathogens is essential for our understanding of the molecular mechanisms underlying the pathogenesis of disease. Identification and functional analysis of virulence genes and their protein products not only furthers our general understanding of mechanisms but can also provide us with specific knowledge that can be used in the development of antimicrobials or vaccines.

What is a virulence factor? A simple definition would be any microbe-specific factor that contributes to the ability of that microbe to cause disease. In reality it is more complex than this. From the perspective of the pathogen a successful infection is defined by multiplication within a host and ultimately transmission. This cycle may or may not involve a negative outcome for the host. In general terms there are several steps in the cycle: 1) attachment to host tissue, 2) growth or invasion at the site of attachment, 3) avoidance and/or resistance to host immune defenses, 4) acquisition of nutrients, 5)

multiplication and finally 6) transmission (Lipsitch and Moxon, 1997). The definition of a virulence factor could be widened to include any microbial-specific factor that contributes to these events. Some factors will be important for microbial cell nutrient acquisition and cell viability in and outside the host. These are essential for the organism's ability survive but may not be specific to the disease process or provoking a host response. There are also factors that are essential for a successful host-pathogen interaction, such as adhesins, which may not themselves be detrimental to the host. One can therefore divide virulence factors into those that are required for microbial cell viability and those which are solely required for disease production.

S. pneumoniae is rarely found outside the human host and requires the host for survival. One might therefore argue that a successful *S. pneumoniae* strain would not benefit from host death but would be better placed living in harmony with its host. A bacterium needs only to have a suitable route for transmission to a second host to ensure continuing survival of the species. Asymptomatic nasopharyngeal carriage by the human therefore could be seen as an ideal. Nasopharyngeal colonization is also a pre-requisite for invasive disease (Bogart *et al.*, 2004). When considering virulence factors in this organism one may wish to include all factors that are important for successful nasopharyngeal carriage as well as those required for disease such as pneumonia or septicaemia.

Information about the mechanisms by which bacteria interact with their host and cause disease has increased dramatically in the post-genomic age. The release of the *S. pneumoniae* genome sequence by TIGR in 1997 was a watershed for pneumococcal pathogenesis research. The organism has a relatively small genome of ~2.1 MB

containing 2236 predicted coding regions. Many virulence factors were known of before the availability of the sequence including the capsule, pneumolysin (Ply), IgA protease, neuraminidase, hyaluronidase, pneumococcal surface adhesin A (PsaA) and a few choline binding proteins (LytA, CbpA and PspA) (AlonsoDeVelasco *et al.*, 1995; McCullers and Tuomanen, 2001b; Mitchell, 2000). A detailed and comprehensive description of all that is known about *S. pneumoniae* virulence factors and disease pathogenesis is beyond the realms of this introduction but some relevant aspects are highlighted in the sections below. Factors involved solely in nutrient acquisition and bacterial viability have been excluded. As discussed later the availability of the genome sequence and genetic screens has accelerated the discovery of additional virulence factors.

1.3.1 Surface associated factors: Role in attachment to host surfaces and internalisation of bacteria

S. pneumoniae attaches to host surfaces in both a carriage state and in disease and attachment is crucial for disease and colonisation. The molecular basis for adhesion to various host surfaces in this organism is complex and has not been fully defined. It has been suggested that adherence and invasion may involve, at very least, a two step process. Initial binding to the host cells leads to cytokine activation. This results in expression of novel receptors on the surface of activated host cells and increased pneumococcal adherence (Jedrzejas, 2001). There is evidence that the receptors and/or structures involved in adherence are different for nasopharyngeal and lung epithelial cells. Several surface structures including proteins have been identified as playing a role in adherence to the upper and/or lower respiratory tract epithelial cells *in vitro* and nasopharyngeal colonisation in animal models.

The cell wall lying below the polysaccharide capsule consists of polysaccharides and teichoic acid and serves as an anchor for cell wall associated surface proteins. Some of these proteins are involved in adhesion to host mucosal surfaces by interactions with specific receptors but in addition they contribute to the hydrophobic and electrostatic surface characteristics of the bacteria, thus facilitating adherence through non-specific physiochemical interactions (Swiatlo *et al.*, 2002). The cell wall also contains phosphorycholine (ChoP) on teichoic acid, which facilitates bacterial interaction with the receptor for platelet-activating factor (rPAF), on human endothelial cells. This interaction with rPAF increases adherence to and invasion of host cells (Cundell *et al.*, 1995b) (Ring *et al.*, 1998) and has been suggested to have a major role in the lung (Rijneveld *et al.*, 2003). The natural ligand for rPAF contains ChoP (Cundell *et al.*, 1995b). The quantity of the surface expressed receptor on host cells is up-regulated by cytokine activation resulting in increased bacterial adherence and internalisation. As well as binding to rPAF, ChoP acts as a docking site for a class of proteins known as choline-binding proteins (CBPs). The major choline-binding protein A (CbpA) has been shown to be an adhesin. *CbpA*- deficient pneumococci have impaired ability to interact with cytokine-activated human epithelial cells, immobilised sialic acid and lacto-N-neotetraose (Rosenow *et al.*, 1997). CbpA has also been demonstrated to interact with polymeric Ig receptor (pIgR) and this interaction has been proposed to be important for the migration of *S. pneumoniae* across the upper respiratory tract mucosal barrier (Lu *et al.*, 2003; Zhang *et al.*, 2000).

Another class of surface-associated proteins are covalently attached to the cell wall via an LPXTG motif, the majority being anchored by the sortase A enzyme. In the R6

genome there are 23 proteins containing the LPXTG motif and only one gene, *srtA*, encoding a sortase. Inactivation of *srtA* decreases adherence of *S. pneumoniae* to human pharyngeal cells *in vitro* suggesting that some of the sortase-target proteins are involved in adherence to and colonization of the nasopharyngeal cells (Kharat and Tomasz, 2003). Work presented in this thesis demonstrates the importance of sortase-anchored cell surface proteins in serotype 4 *S. pneumoniae* in adhesion to host cells. IgA protease has a characteristic LPXTG motif consistent with its covalent attachment to the cell wall by a sortase enzyme. One function of the enzyme in adhesion has been elucidated by Weiser et al (Weiser *et al.*, 2003). They showed increased adherence of *S. pneumoniae* to lung epithelial cells in the presence of human IgA. The effect required functional IgA protease and is thought to be brought about by cleavage of opsonising IgA by the protease, resulting in a change in surface charge and increased physical proximity of the pneumococcal cell-wall ChoP to rPAF.

Another sortase-anchored enzyme in *S. pneumoniae*, neuraminidase or NanA, acts at the bacterial/host cell interface and facilitates *S. pneumoniae* adherence and colonization. The enzyme cleaves sialic-acid from host cell glycolipids, glycoproteins and oligosaccharides and is thus thought to expose N-acetyl-glycosamine and possibly other receptors for bacterial adherence (Camara *et al.*, 1994; Krivan *et al.*, 1988). Both *Neisseria meningitidis* and *Haemophilus influenzae* are capable of mimicking host structures by decorating their lipopolysaccharides with sialic acid and may use this as a protective mechanism against host defenses. These organisms are thought to compete with *S. pneumoniae* for nasopharyngeal colonisation. NanA expressed by *S. pneumoniae* has been shown to desialylate the cell surfaces of both these species *in vitro* (Shakhnovich *et al.*, 2002). Interfering with the biology of these potential

competitors in this way may be a mechanism by which *S. pneumoniae* can gain a selective advantage for colonization of the host upper respiratory tract.

Many Gram-positive bacteria utilize extracellular matrix (ECM) molecules to mediate attachment to host surfaces (Chhatwal, 2002). There are many examples of surface exposed proteins that facilitate interactions with matrix molecules such as fibronectin, collagen and fibrinogen in *Staphylococcus, Enterococcus* and other streptococcal species (Foster and Hook, 1998; Joh *et al.*, 1999; Patti *et al.*, 1994; Rozdzinski *et al.*, 2001) but *S. pneumoniae* has been notably absent in these reports. *S. pneumoniae* was shown to bind to immobilized fibronectin in a dose dependent manner and this binding is reduced by trypsin treatment of bacteria suggesting a surface exposed protein-dependent mechanism (van der Flier *et al.*, 1995). In these studies the level of binding varied in different clinical isolates and strains of different serotypes. Isolates of types 7F, 14, 18C and 19F bound most efficiently whereas isolates of serotype 2 and 4 bound poorly suggesting that either capsule composition effects fibrinogen interactions or that fibrinogen binding is not a universal mechanism of attachment for *S. pneumoniae* and not all strains habour fibrinogen-binding proteins. However a more recent study characterised a surface associated pneumococcal protein, PavA, which binds fibronectin and is associated with virulence. The *pavA* gene has sequence similarity to Fbp54, a fibronectin binding protein in *S. pyogenes,* and was found to be present in 64 independent isolates tested representing different serotypes including serotypes 2 and 4. The capsule was shown to have an inhibitory role in attachment to fibronectin in this study.

PsaA was originally reported as a surface adhesin in *S. pneumoniae* based on sequence similarity to other streptococcal adhesins (Sampson *et al.*, 1994). *psaA*- mutants are virtually avirulent in animal models of disease and show reduced ability to adhere to type II pneumocytes *in vitro* (Berry and Paton, 1996). The *psaA* gene is part of an operon consisting of three genes which encode a manganese transport system (Dintilhac *et al.*, 1997). It is not clear whether the PsaA protein is itself the adhesin or whether the decrease in adhesion seen in *psaA*-mutants is a reflection of the requirement for manganese as a regulator of adhesin expression.

Regulation of expression of these various surface factors and bacterial adherence to mucosal surfaces is not fully understood though one mechanism has been demonstrated. Phase variation has been shown to affect the cell wall composition, surface protein and capsule polysaccharide expression. This phenotypic variation is also associated with differences in colonization and invasion capacities in strains (Weiser *et al.*, 1996). Transparent phase variants, which have less capsular polysaccharide, more teichoic acid and more CbpA, than their opaque counterparts, show greater adherence than opaque variants (Kim and Weiser, 1998; Weiser and Kapoor, 1999).

1.3.2 Factors involved in the avoidance and/or modulation of host immune defenses

S. pneumoniae must evade host immune responses during colonization and invasive disease to survive. Some of the factors involved and mechanisms by which the bacteria achieves this have been elucidated. Often these factors have additional roles involved in virulence in other ways.

The polysaccharide capsule has long been recognized as playing a major role in virulence and is arguably the most important virulence factor in *S. pneumoniae* by preventing bacterial phagocytosis (McCullers and Tuomanen, 2001a). Acapsular isolates rarely cause invasive disease in humans and are greatly attenuated in virulence in animal models of infection (Malley *et al.*, 2001). Reduced capsular expression results in greater access of antibodies and complement to the bacterial surface and hence increased clearance by the host immune system (Magee and Yother, 2001). Capsular polysaccharides are highly immunogenic and antibodies against them protect against invasive disease with the homologous serotype by induction of opsonophagocytosis.

The CBP, pneumococcal surface protein A or PspA, is common to every *S. pneumoniae* strain studied and is required for virulence (Crain *et al.*, 1990; McDaniel *et al.*, 1987). It has been shown to inhibit complement binding and activation thus being important in reducing complement mediated clearance and phagocytosis of bacteria (Briles *et al.*, 1997; Tu *et al.*, 1999) (Ren *et al.*, 2004). It is also a lactoferrin–binding protein and as such may have an additional role in nutrient acquisition during infection (Hammerschmidt *et al.*, 1999).

Pneumolysin, a thiol-activated, multifunctional cytolysin released by *S. pneumoniae*, is another well-characterized pneumococcal virulence factor. It is a cytoplasmic protein which is released on bacterial cell autolysis and also during bacterial growth (Balachandran *et al.*, 2001). The role of the toxin in virulence is complex but it is important both in early and later stages of disease. It appears to have many independent functions that contribute to the pathogenesis of *S. pneumoniae* disease, including affects on host response, tissue inflammation and even enhancement of bacteria adhesion to

host cells. It interacts with cholesterol and inserts into host cell membranes resulting in pore formation and cell lysis. It is also capable of a variety of detrimental modulatory effects on the host immune system including directly activating the classical complement pathway in the absence of antibody resulting in a reduction in serum opsonic activity (Mitchell *et al.*, 1991; Paton *et al.*, 1984). The enzyme can directly inhibit human neutrophil and monocyte respiratory bursts, chemotaxis, bacteriocidal activity and production of lymphokines and immunoglobulins (Paton and Ferrante, 1983; Rubins and Janoff, 1998). The use of cDNA microarrays has enabled the identification of 142 genes in THP-1 human monocyte cells, which are responsive to *S. pneumoniae* in a pneumolysin-dependent manner and some of which may be involved in mediating these affects (Rogers *et al.*, 2003). Infection of mice with pneumolysin-deficient mutants results in chronic bacteraemia rather than the rapidly progressive sepsis observed with isogenic wild-type strains. This is consistent with the hypothesis that the presence of pneumolysin impairs host resistance to overwhelming infection and ability to prevent exponential bacterial growth (Benton *et al.*, 1995; Benton *et al.*, 1998).

1.3.3 Virulence factors involved in damage to host tissue and bacterial spread

Development of inflammation, damage and destruction to host tissues is central to the pathophysiology of disease. The induction of inflammation also increases bacterial invasion into respiratory epithelium and vascular endothelium, at least in part, by up-regulation of receptors on host cells and exposure of other host components required for bacterial adherence. The primary inflammatory stimulant is the *S. pneumoniae* cell wall. (Tuomanen *et al.*, 1985). Binding of the peptidoglycan to CD14, a host cell surface receptor, and toll-like receptor 2 initiates a cytokine cascade. Other cell wall

components can also trigger cytokine release by CD14 independent pathways (Cauwels *et al.*, 1997). The result is the induction of NF-κ B, TNFα and various interleukins (Bergeron *et al.*, 1998; Spellerberg *et al.*, 1996).

Pneumolysin is also an important inflammatory stimulant. Pneumolysin-deficient *S. pneumoniae* strains exhibit reduced virulence in mice with delayed development of the cellular inflammatory response (Kadioglu *et al.*, 2000). The toxin's cytotoxic activity results in host cell lysis and tissue destruction. Pneumolysin's other activities include inhibition of ciliary beating on the human respiratory epithelium, which could result in impaired bacterial clearance from the respiratory tract (Feldman *et al.*, 1990; Feldman *et al.*, 1991) and separation of epithelial cell tight junctions (Rayner *et al.*, 1995), that may expose host receptors for binding of *S. pneumoniae* and enhance bacterial invasion and spread.

Hyaluronate lyase (Hyl), which degrades hyluronic acid, a component of connective tissue, has also been implicated in destruction of tissue integrity and allowing *S. pneumoniae* greater access to deeper tissues and the bloodstream (Berry and Paton, 2000; Mitchell, 2000). Lastly, two surface exposed plasminogen-binding proteins, α-Enolase (Eno), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) have been described in *S. pneumoniae* and are proposed to have a role in promoting penetration of bacteria through the basement membranes (Bergmann *et al.*, 2001; Bergmann *et al.*, 2004). Plasminogen is a monomeric proenzyme of the serine protease plasmin and plays a crucial role in fibrinolysis and the degradation of ECM (Vassalli *et al.*, 1991). Many pathogens capture plasmin(ogen), potentially allowing the bacteria to acquire surface

associated proteolytic activity. This activity has been shown to facilitate their invasion and dissemination in the infected host (Lottenberg *et al.*, 1994; Lottenberg, 1997).

1.4 CONTEMPORARY STRATEGIES FOR IDENTIFICATION OF VIRULENCE GENES AND *IN VIVO* EXPRESSED GENES AND THEIR APPPLICATION TO *S. PNEUMONIAE*

More classical approaches to virulence gene identification in general and in *S. pneumoniae* that led to the discovery of many of the factors described above will not be discussed here but have been detailed elsewhere (Camilli *et al.*, 2001). I will focus here on more contemporary strategies relating to the thesis and specifically those used to identify *in vivo* expressed genes.

1.4.1 Sequence scanning and genomic analysis

At the beginning of 2004 there were over 100 complete genome sequences open to public access via the TIGR or other institutional websites or available on a more limited basis. Coupled with the availability of computer resources including numerous algorithms this has allowed the comparison of whole proteins, selected protein motifs, genes or nucleotide sequences across species and genera enabling the identification of new candidate virulence genes or vaccine targets. In this way orthologues of proteins known to be important for virulence in one organism can be identified in another and assigned a putative function. Regions within genomes that have a different G+C content can be highlighted. These often indicate segments of DNA that have been acquired from another species by horizontal transfer and often contain virulence genes. This type of work now forms a basis for identifying new virulence factors and providing some knowledge of possible function prior to further in-depth molecular studies.

1.4.2 Functional genomics in *S. pneumoniae*

Access to the complete pneumococcal genome sequence has allowed systematic and comprehensive searches for virulence factors in particular categories and has aided the identification of hundreds of potential new vaccine candidates. For example, *S. pneumoniae* produce a family of surface proteins that attach to the cell wall teichoic acid or the membrane-associated lipoteichoic acid via a choline-binding repeat domain consisting of 2-10 sequence repeats (Garcia *et al.*, 1998). CbpA, the largest and the most abundant of the choline-binding proteins, functions as a surface adhesin and plays an important role in nasopharyngeal colonization (Rosenow *et al.*, 1997). By performing genome-wide searches with the C-terminal choline binding region of CbpA 12 further genes encoding choline-binding motifs were identified (Gosink *et al.*, 2000). Construction of mutants was then used to demonstrate that several of these choline-binding proteins are involved in adherence to epithelial cells, nasopharyngeal colonization and/or sepsis (Gosink *et al.*, 2000). Another group identified surface-exposed proteins as potential vaccine candidates by examining the genome sequence for open reading frames (ORFs) with motifs associated with transport, cell-wall anchorage, or choline-binding as well as ORFs with similarity to known virulence factors in other bacteria (Wizemann *et al.*, 2001). They discovered that genes encoding putative surface exposed proteins represent 4% of the *S. pneumoniae* genome. They expressed approximately 100 of these in *E. coli* enabling protein purification for testing as immunogens in mice. Five proteins were shown to be protective immunogens in a mouse sepsis model involving intra-peritoneal inoculation of bacteria (Wizemann *et al.*, 2001).

The ability of *S. pneumoniae* to sense and adapt to changing host environments is likely to be important for bacterial survival and the pathogenic process. This adaptation is often mediated in bacteria by two-component signal transduction systems (TCSTs) comprising a sensor histidine kinase and a cognate response regulator (Stock *et al.*, 2000). The use of similar genomic scanning in two independent surveys has identified 13 two-component signal transduction systems (TCSTs) in the *S. pneumoniae* genome. Mutations were constructed in genes encoding either both or one of the components of the TCSTs (Lange *et al.*, 1999; Throup *et al.*, 2000). The majority of these systems are required for virulence in a murine model of pneumonia, though as discussed in more detail later there was disconcordance in the findings in the two studies.

1.4.3 Genetic Screens: Signature tagged mutagenesis

Since the mid 1990s, at the same time as the first genome sequences were becoming available, there have been numerous publications on genetic screens to identify new virulence factors. All methods have drawbacks and ideally one would use more than one approach to complement findings. Signature Tagged Mutagenesis (STM) screens aim to identify factors that are essential for infection by negative selection (Hensel *et al.*, 1995) and have now been applied to a myriad of pathogens (reviewed in (Chiang *et al.*, 1999; Mecsas, 2002)). In most cases STM utilizes transposon mutagenesis to generate insertion mutants, though other methods of mutagenesis have been used. A select number of differentially marked (signature tagged) transposons are employed, such that each can be identified in a larger pool. These signature tags consist of a randomized sequence of around 40 bp, that is unique to each specific tag, flanked by sequence arms common to all tags. The common arm sequence allows recognition of any tag by a set of universal primers and the variable region allows identification of

each specific tag. Individual mutants with different tags are grown separately, pooled and inoculated into an animal host or cell culture system. This is termed the input pool. After a period of time bacteria are isolated from the animal (output pool) and the presence or absence of mutants within the output pool is detected by PCR amplification of the tags and hybridization to a dot-blot containing all of the unique tag sequences. Loss of a mutant from input to output is predicted to be due to disruption of a gene required for bacterial survival and/or multiplication *in vivo*. The disrupted site can be determined by analysis of the mutant. STM screens have allowed the simultaneous identification of multiple factors in a single infection and as such has been an incredibly powerful tool for mass identification of virulence factors. It is not however without limitations. Firstly, since the input contains a mixed population theoretically there could be trans- complementation of a mutation in one bacteria by a product from a second bacteria, eg. toxin production. This would lead to failure of selection of that mutation as disrupting a virulence gene. Secondly, mutants must be viable *in vitro*, genes that are essential for growth during the selection process will not be obtained, although this may not necessarily be seen as a disadvantage. Thirdly, the use of transposon insertion as the method of mutagenesis can lead to the generation of polar effects on neighbouring genes, which can complicate the interpretation of the output data. Lastly STM as with most other genetic screens will not identify virulence factors that are functionally redundant with other bacterial proteins, as a single mutation can be functionally complemented.

1.4.4 STM and *S. pneumoniae*

There have now been three published STM screens in *S. pneumoniae*, each in a different serotype strain; 3 (Lau *et al.*, 2001), 4 (Hava and Camilli, 2002) and 19F (Polissi *et al.*,

1998). Each screen used the murine model of pneumonia to identify genes that are essential for bacterial survival. All three screens then also included further analysis of mutants in either a bacteraemia model or a model of nasopharyngeal carriage. Between the three screens over 200 new virulence genes were identified. All screens identified some of the already known virulence genes, and genes whose products were predicted to be involved in cell adhesion, metabolism, transportation across the cell membrane and regulation to name a few. The largest group in each screen were hypothetical genes with unknown function. Interestingly there is remarkably little overlap in the genes identified by each screen, giving some insight into the virulence properties of different serotypes. However it is difficult to make meaningful comparisons when the screens did not approach saturating levels and when publications did not reveal the identity of all the genes hit.

A difference in the results from these STM screens and those reported in other gram-positive pathogens is the percentage of mutagenised strains identified as attenuated *in vivo*. Screens in *S. pneumoniae* have found that up to 20% of the mutated strains screened are decreased in virulence as compared to a figure of nearer 7% noted in STM screens performed in other pathogens (Autret *et al.*, 2001; Jones *et al.*, 2000; Mei *et al.*, 1997). The finding that such a high proportion of mutants are attenuated may not be all that surprising given that *S. pneumoniae* has a relatively small genome and no reservoir outside the human host. One would expect that, to avoid redundancy, a large percentage of its genes would be essential for colonization and survival in the human host. Many of these genes required for successful interaction with the mucosal surface of the nasopharynx leading to colonisation may also be involved in more detrimental interactions with the upper and lower respiratory tract such as pneumonia.

1.4.5 Genetic Screens: *in vivo* expression technology

In vivo expression technology (IVET) was originally conceived upon the premise that most virulence factors are transcriptionally induced at some stage during infection (Mahan *et al.*, 1993). Although certain environmental parameters can be mimicked *in vitro* often the signals required for virulence gene induction cannot be guessed at and reproduced *in vitro* and a subset of genes will only be expressed *in vivo*. An advantage of IVET is that the live host and infection are used identify *in vivo* induced (*ivi*) genes. There are several variations of IVET but all rely on the generation of transcriptional fusions of genomic sequences to a reporter gene encoding enzymatic activity (Angelichio and Camilli, 2002). The variation lies in the actual type of reporter gene used. The original IVET took advantage of the fact that a purine auxotroph (Δ*purA*) of *Salmonella enterica* serovar typhimurium is rapidly eliminated from the mouse unless it is complemented in trans (Mahan *et al.*, 1993). *Ivi* genes were identified by ligating random genomic fragments upstream of a promoterless *purA-lacZY* synthetic operon present on a suicide vector. This library was transferred into *S. typhimurium* and the resultant mutants were used to infect mice. Growth in the mouse provided positive selection for genes that expressed an *ivi* gene driving the expression of *purA-lacZY* as these strains became prototrophic and survived. Other variations of IVET employ antibiotic resistance genes as selectable markers, pore-forming listeriolysin or a recombination based system (RIVET discussed later) (Camilli *et al.*, 1994; Gahan and Hill, 2000; Mahan *et al.*, 1995). These IVET systems have now been used to identify *ivi* genes in numerous pathogens (Angelichio and Camilli, 2002). But like STM, IVET is not without disadvantages. The screen will select for any *ivi* genes. The genes identified may not necessarily be required for virulence and, indeed, not truly encode

virulence factors. This is illustrated in the original IVET screen. In taking bacteria from growth in rich medium *in vitro* to the more restricted nutritional environment of the mouse peritoneum one would predict induction of expression of numerous nutrient acquisition, metabolic and catabolic genes, which was what the initial IVET detected. Another drawback with IVET is that selection *in vitro* for the input pool may require the transcriptional fusion to be silent. Thus genes that are transcribed *in vitro* will be excluded from the input even if are further induced *in vivo* and play a pivotal role in infection.

1.4.6 *In vivo* gene expression in *S. pneumoniae*

Classical IVET has not as yet been reported as being used in *S. pneumoniae* but other techniques have been employed to examine gene induction *in vivo*.

Differential fluorescence induction technology (DFI) has been used to identify promoters of *S. pneumoniae* that are induced under various *in vitro* and *in vivo* conditions. The technology relies on a green fluorescent protein (GFP) reporter gene to indicate expression from a given promoter under test conditions (Valdivia and Falkow, 1997). Random small fragments of chromosomal DNA from *S. pneumoniae* were ligated upstream of a promoterless *gfp* gene on a shuttle plasmid. The resulting library was transformed into *S. pneumoniae* and subjected to *in vitro* conditions such as high osmolarity, temperature shift and change in CO_2 concentration and also *in vivo* screening in models of otitis media, pneumonia and growth in an intraperitoneal chamber implant (Marra *et al.*, 2002a; Marra *et al.*, 2002b). After a period of growth *in vitro* or time in the animal model, fluorescent-activated cell sorting (FACS) was used to sort harvested bacteria based on their fluorescent levels. The roles in virulence of genes

identified as being induced *in vitro* or *in vivo* were ascertained by construction of deletion mutants and subsequent testing *in vivo*. Several genes were found to be induced twofold or greater in the selected *in vitro* or *in vivo* conditions. Of these 25 were mutated and14 of these were attenuated in one or more of the animal models (Marra *et al.*, 2002a). What exact roles they play is yet to be elucidated.

Two other studies have looked specifically for evidence of differential virulence gene expression *in vitro* and *in vivo*. The first compared mRNA levels for specific virulence factors in total RNA harvested from serotype 3 *S. pneumoniae* grown in dialysis bag implanted into the peritoneal cavity of mice to those from control *in vitro* cultures (Orihuela *et al.*, 2000). However the environment encountered by bacteria in a peritoneal dialysis bag cannot simulate that which occurs in a natural infection. Levels of mRNA for 4 genes (*ply*, encoding pneumolysin, *capA3*, encoding a gene involved in capsule biosynthesis, *pspA*, encoding a surface antigen involved in iron acquisition and *lytA*, the major autolysin) were examined by northern blotting. Only *ply* and *capA3* showed changes in expression *in vivo*. Both were up-regulated 2-3 fold as compared to levels in *in vitro* grown bacteria. There was no difference in expression of *lytA* and *pspA* *in vivo* compared to *in vitro*. The second study compared the relative abundance of mRNA transcripts of five genes in serotype 2 *S. pneumoniae* harvested from the blood of mice at 12 and 24 hours following intra-peritoneal infection to mRNA levels present in organisms grown in serum broth (Ogunniyi *et al.*, 2002). The group exmained *ply*, *pspA*, *psaA*, encoding pneumococcal surface antigen, *cbpA*, encoding choline-binding protein A, and the capsule biosynthetic gene *cps2A*. The expression of all but *cbpA* were found to be up-regulated during infection. These types of studies provide some direct molecular evidence that known virulence-associated genes of *S. pneumoniae* are

differentially expressed *in vivo,* but we are a long way from being able to pinpoint the timing of and identify the signals for their expression.

1.4.7 DNA microarrays and comparative genomics

The main advantage of microarrays over other systems is they have provided us with the ability to measure simultaneously the presence of tens of thousands of different nucleic acid sequences, thus allowing the analysis of gene expression and ability to perform comparative genomics on a genome-wide scale. The method takes advantage of the fact that, due to the specificity and affinity of complementary base pairing, a DNA copy of a gene can provide an ideal reagent for specific and quantitative detection of that sequence in a mixture of others. The array is essentially a microscopic chequerboard representing thousands of different such DNA sequences.

There are several methods of producing these microarrays. Commercially produced arrays are usually high-density oligonucleotide arrays constructed by synthesizing short (~25-mer) oligonucleotides *in situ* on glass wafers. The technology required for production makes these types of arrays too expensive for 'in house' production, which tend to be generated by spotting/printing single or double stranded DNA onto glass slides. Here the DNA used is either cloned, synthesised or polymerase chain reaction-amplified products. The relative abundance of each or any of these gene sequences in two DNA or RNA samples can be measured. The two samples are differentially labeled using fluorescent dyes. They are then mixed and hybridised with the arrayed DNA spots. However if RNA samples are to be assessed the initial step before labeling and hybridization is cDNA synthesis using RNA as the template. This cDNA is taken forward to subsequent steps. After hybridization, measurements are taken for each

fluorescent dye separately for each spot. These are used to determine the ratio, and in turn the relative abundance, of the sequence of each specific gene in the two original samples.

As well as providing data on quantitative differences between samples enabling relative abundance of mRNAs to be determined, arrays can be used to determine absolute differences between strains allowing comparative genomics. It has become increasingly apparent that within many single bacterial species there is enormous variability between genomes of different strains. Microarray analysis has become a well-established tool for exploring the distribution of genes among collections of isolates from the same species and identifying virulence determinants. For example comparative genomics on 15 clinical *H. pylori* clinical isolates identified a set of 1,281 core genes (out of 1,643) common to all strains (Salama *et al.*, 2000). Amongst those variable genes was a strain specific pathogenicity island that encodes a type IV secretion system and an additional set of genes that appeared to be co-inherited with this island that encoded new putative virulence factors. One important aspect of comparative genomics is that the information obtained is based on the microarray used. If the tester strain has additional unique sequences these will not be detected.

As previously mentioned it is clear from epidemiological, molecular and animal studies that all strains of *S. pneumoniae* are not equally efficient at colonizing a host or causing disease. These differences are probably due to a combination of things; capsular make-up, specific surface protein expression and as yet other uncharacterized genetic differences. Comparison of the completed or partially completed genome sequences of 4 different serotypes clearly reveals the substantial genetic diversity between serotypes

(Hoskins *et al.*, 2001; Tettelin *et al.*, 2001). The extent of diversity also varies with specific populations. The genetic diversity as assessed by MLST of the invasive pneumococcal population irrespective of the serotype is significantly less than the carriage population (Brueggemann *et al.*, 2003). Microarrays have been used successfully to perform comparative genomics and determine genetic differences between strains of *S. pneumoniae* and their use in more extensive comparative studies between clinical invasive and carriage isolates will no doubt play a significant part in adding to our understanding of which factors contribute to virulence.

1.4.8 Global expression studies

A central requirement for all pathogenic organisms is to possess the ability to adapt to changing environmental conditions. This adaptation must involve regulating gene expression in response to environmental signals. There has been a natural assumption that there is up-regulation of expression of genes that are needed in a particular environment and down-regulation of those that are not needed under the same conditions. Similarly one might expect the same principle to apply to virulence genes, ensuring their expression in the appropriate host environment in which they are required and this hypothesis has been the basis of array studies attempting to further identify virulence factors. This simple assumption, though, may not hold true. Two studies in *Saccharomyces cerevisiae* showed little overlap between genes identified as having a change in mRNA expression in specific conditions, as determined by microarray profiling, and genes necessary for survival in the same conditions. The conditions looked at included exposure to DNA-damaging agents, change in carbon source or challenge with an anti-fungal compound (Birrell *et al.*, 2002; Giaever *et al.*, 2002). This indicates that a significant proportion of genes identified by expression profiling may

not be absolutely required for an adaptive response, making microarray studies without further experimental work unhelpful. Some form of genetic analysis is necessary to identify the subset of genes which are required for adaptation to the condition in question.

Bearing this caveat in mind bacterial expression patterns under different *in vitro* conditions that mimic particular host conditions have been examined with the use of microarrays (Merrell *et al.*, 2003; Tao *et al.*, 1999) with the aim of identifying specific genes that are up-regulated in specific conditions. Changes in gene expression compared to wild-type have also been determined for bacterial mutants lacking putative virulence gene regulators in attempts to identify their targets (Kato-Maeda *et al.*, 2001; Schoolnik, 2002a, b). In this way multiple new virulence factors have been identified and the conditions required for their expression elucidated. Finally eucaryotic microarrays have also been used to study the host response to pathogen interactions (Boldrick *et al.*, 2002).

Microarrays have been used successfully to examine gene expression on a genome-wide basis in *S. pneumoniae*. To date there have been over a dozen studies published that utilize microarray technology to examine *S. pneumoniae* gene expression profiles with the aim of determining proteins that are co-expressed in specific growth conditions or in mutant backgrounds. These have primarily revolved around further characterizing genes involved in competence or identifying regulons of TCSTS. They will be discussed in the section on regulatory networks in *S. pneumoniae*.

As with all the methods that have been mentioned microarray analysis too has its limitations. There has been a lag between the initial publication of microarray studies and production of standards in both experimental and statistical methodology allowing meaningful interpretation of data and comparison of studies. There are concerns about the relative stability of different mRNA species when wishing to assess global expression changes (Bernstein *et al.*, 2002). The choice of *in vitro* growth conditions will always affect the expression patterns that are obtained and is unlikely to match the exact conditions experienced *in vivo*. Ideally one would like to obtain nucleic acid from bacteria from host infected tissue but a major technical hurdle to date has been to harvest and purify adequate quantities of bacterial RNA to confidently monitor bacterial gene expression at different stages of and in different types of infection. Recently the gene expression patterns of *Vibrio cholerae*, recovered from the stool of human patients has been examined (Merrell *et al.*, 2002) and expression studies have been performed on RNA from bacteria obtained from blood in septicaemic murine infections. Despite drawbacks microarray technology has and will continue to provide a wealth of data on virulence genes and their expression.

1.5 VIRULENCE GENE REGULATION AND REGULATORY NETWORKS IN *S. PNEUMONIAE*

During the process of survival and multiplication in the host *S. pneumoniae* will experience different microenvironments. Bacteria need to be able to adapt to the changing environment and focus their cellular processes to efficient functioning in new environments. For example, transition from a state of colonization to invasive disease involves adapting from survival at a mucosal surface to survival within host tissue and the bloodstream. This inevitably involves the control of expression of various factors

involved in housekeeping and metabolic processes but also virulence factors. Research focusing on regulation of gene expression in *S. pneumoniae* and more specifically elucidation of regulatory networks within the organism can be grossly divided into two; (1) Work detailing the complex networks involved in development of competence and transformation and (2) Studies investigating other regulatory networks, primarily focusing on elements of virulence. It is interesting that, historically, studies into competence have dominated research on gene regulation in *S. pneumoniae*. At the time of starting this thesis project in January 2000 virtually nothing was published on virulence gene regulation in *S. pneumoniae*. It has only been relatively recently, perhaps sparked by the identification of genes with regulatory function as well as access to the genome sequence, that there has been much interest specifically in virulence gene regulation in this organism. Most of that discussed below has come to light during the time of my project. The regulatory cascade involved in development of competence has been touched on early in this work and reviewed elsewhere (Morrison and Lee, 2000). Since competence is not the primary focus of this thesis it will not be discussed in further detail except to comment on its links into other regulatory systems and virulence.

1.5.1 Transcriptional regulators

The three STM screens not only identified many new virulence genes but also indicated that tissue-specific virulence factors might exist. The presence of tissue specificity in virulence factor requirement may also indicate the presence of mechanisms to regulate their expression and/or function in different host tissues. In one STM study the mutant pools were assessed simultaneously for survival in models of lung infection and septicaemia (Lau *et al.*, 2001). A third of the mutants were found to be essential for

both types of infection, a third each were essential for either pneumonia or septicaemia. In another STM study the phenotypes of a subset of the total 387 attenuated mutants were examined in two other murine models, namely bacteraemia and nasopharyngeal carriage. This identified four classes of mutants defective in infection models of the: lung, lung and blood, lung and nasopharynx, and all three tissues (Hava and Camilli, 2002). These findings indicate that some factors function in multiple environments whereas others only in specific environments though what external triggers are responsible for altering their expression and activity is not clear.

In total 20 different putative or known transcriptional regulators (excluding the TCSTS) were identified by the three STM screens. Some have now been assigned direct or indirect regulatory roles (Hava *et al.*, 2003b). Possible regulatory roles have been predicted for the majority of the remaining factors (Table 1.1).

Transcriptional regulators other than those identified in the STM screens have been investigated for their role in virulence gene regulation. A screen to isolate mutants that affected expression of the competence related genes *comCDE* lead to the identification of the transcriptional repressor CstR (Chastanet *et al.*, 2001). The repressor was shown to also repress expression of Clp ATP-dependant proteases, one of which had already been identified as a virulence factor by STM. RegR, a LacI/GalR family regulator, modulates virulence and competence in *S. pneumoniae* (Chapuy-Regaud *et al.*, 2003). It represses expression of hyaluronidase, a cell wall anchored virulence factor (Berry *et al.*, 1994; Berry and Paton, 2000; Paton *et al.*, 1993) *in vitro* and also acts upstream of the CiaRH TCSTS to activate competence. Its activity is sensitive to alterations in pH. RegM, a homologue of the staphylococcal catabolite repressor protein, CcpA, is

required for virulence in a murine model of septicaemia. It does not appear to be involved in regulation of expression of virulence factors CbpA, PspA, PsaA and Ply but was involved in regulating expression of the capsular polysaccharide biosynthesis locus (*cps*) (Giammarinaro and Paton, 2002) and as such is the first gene implicated in the transcriptional regulation of capsule genes. The function of neither RegM or RegR is likely to be solely the control of virulence gene expression as both have similarity to regulators that control metabolic pathways in other organisms.

Some of these regulators may be involved in defining the tissue specific expression of the virulence genes identified in the STM screens. Finding their targets however is simpler than determining the signals to which they respond but without this knowledge we cannot begin to dissect the regulatory networks that are important in a switch of phenotype from a relatively benign one seen in carriage to an aggressive one seen in invasive disease. What is becoming clear though is that there are complex regulatory networks that link the expression of virulence genes to other cellular pathways including competence.

Table 1.1 Transcriptional regulators identified by STM

TIGR4 locus[a]	Gene	Serotype[b]	Regulatory role[c]	Reference
SP0141		4	Unknown	(Hava and Camilli, 2002)
SP0246		4	Unknown, DeoR family	(Hava and Camilli, 2002)
SP0247		4	Unknown	(Hava and Camilli, 2002)
SP0306	*smrD*	3,4	Putative cellobiose metabolism	(Hava and Camilli, 2002; Lau *et al.*, 2001)
SP0461	*rlrA*	4	Regulation of *rlrA* islet genes	(Hava *et al.*, 2003a)
SP0807		4	Regulation of septation ring formation	(Hava and Camilli, 2002)
SP0927	*smrC*	3,4	Regulator of *cbpE* operon (putative)	(Hava and Camilli, 2002; Lau *et al.*, 2001)
SP0975	*smrF*	3	mRNA stabilization (putative)	(Lau *et al.*, 2001)
SP1115		4	MutR regulator (putative)	(Hava and Camilli, 2002)
SP1278	*pyrR*	4	Pyrimidine metabolism	(Hava and Camilli, 2002)
SP1433		4	Unknown, AraC family	(Hava and Camilli, 2002)
SP1638	*smrB*	3	Iron dependant regulator (Putative)	(Lau *et al.*, 2001)
SP1800	*mgrA*	4	Virulence gene regulator (Putative), *mga*-like	This work
SP1830		4	Regulation of phosphate transport (putative)	(Hava and Camilli, 2002)
SP1854	*galR*	4	Galactose operon repressor	(Hava and Camilli, 2002)
SP1856		4	Unknown, MerR family	(Hava and Camilli, 2002)
SP2131		4	Pentitol metabolism (putative)	(Hava and Camilli, 2002)
SP2142		4	Unknown, ROK family	(Hava and Camilli, 2002)
SP2195	*cstR*	3	Regulator of Clp proteases	(Chastanet *et al.*, 2001; Lau *et al.*, 2001)
Unclear[d]		19F	Sugar metabolism (putative)	(Polissi *et al.*, 1998)

[a]The TIGR4 annotated gene name. [b]The serotype in which each regulator was identified by STM. [c]Putative regulatory targets based on sequence similarity and information or previous studies. [d]No sequence information or locus details available.

1.5.2 Two component signal transduction systems (TCSTS)

TCSTS are found in a wide range of bacteria where they have been shown to regulate diverse processes, including chemotaxis, nutrient utilization, surface adhesion, the switch between aerobic and anaerobic metabolism and virulence gene expression (Stock *et al.*, 2000). TCSTSs consist of a sensor kinase that autophosphorylates in response to an environmental stimulus and a cognate response regulator (RR) to which an active phosphate is transferred. Phosphorylation of the RR results in a downstream response. The phosphorylated RR often acts as a DNA-binding protein, binding to promoter regions resulting in changes in gene expression. There are genes encoding 13 TCSTSs in the *S. pneumoniae* genome, many of which were also identified in the STM screens as being required for disease. The role of each in virulence has been assessed but the specific function of only 6 of them have been addressed in any detail (Table 1.2).

Table 1.2 The roles of TCSTS in *S. pneumoniae*

Gene names	Known or putative regulatory role	Reference
CiaRH	Regulation of competence, induction of *htrA*	(Sebert *et al.*, 2002)
Com DE	Quorum sensing, regulation of competence and virulence genes	(Peterson *et al.*, 2000)
BlpHR	Production of bacteriocin-like proteins	(de Saizieu *et al.*, 2000)
MicAB	Competence repression, oxygen/redox potential sensing	(Echenique and Trombe, 2001a; Kadioglu *et al.*, 2003)
VncRS	Induction of *vex123* transcription. Possible role in vancomycin tolerance	(Novak *et al.*, 1999b; Robertson *et al.*, 2002)
PnpRS	Regulation of *psaA*, *psaB*, *psaC* and resistance to oxidative stress	(McCluskey *et al.*, 2004; Novak *et al.*, 1999a)

Three published studies examine the involvement of TCSTS in disease in murine models of infection. Lange and colleagues created knockout mutations in genes encoding components of eleven TCSTS of serotype 3 and serotype 22 *S. pneumoniae* strains and found no difference in virulence between the mutants and wild-type strains in a murine septicaemia model (Lange *et al.*, 1999). In contrast deletions in the genes encoding components of eight of the same TCSTS in serotype 3 *S. pneumoniae* resulted in attenuation of strains in a murine model of pneumonia (Throup *et al.*, 2000). The route of inoculation, therefore, appears to be crucial for the role of the TCSTS. Intraperitoneal inoculation circumvents any requirement for adaptation at the mucosal surface of the respiratory tract and this may explain why TCSTS mutants show different phenotypes with respect to ability to cause disease in a septicaemia versus a pneumonia model. These two studies were complimented by examination of the role of the 8 TCSTS, in which mutants were attenuated for pneumonia, in persistence in nasopharyngeal colonization (Sebert *et al.*, 2002). It was found that only the CiaRH TCSTS was required for efficient nasopharyngeal carriage in a murine model. Microarray analysis comparing expression profiles of a CiaRH deficient mutant to wild-type bacteria identified 46 genes that were differentially expressed in the two strains and were thus potentially regulated by the system. The authors hypothesized that the defect in the colonizing ability of the Δ*ciaRH* strain may have been at least in part mediated by a reduction in the product of *htrA*, a serine protease. This gene was down-regulated 37 fold in the Δ*ciaRH* strain, and was shown to be required for colonization (Sebert *et al.*, 2002). The CiaRH system is also a strong repressor of competence development (Echenique *et al.*, 2000; Guenzi *et al.*, 1994).

There are at least two quorum-sensing systems in *S. pneumoniae* each containing a TCSTS. Both have been implicated in virulence (Lau *et al.*, 2001; Throup *et al.*, 2000) which is highly suggestive that cross-talk between bacteria plays a role in *S. pneumoniae* pathogenesis. The well characterised *com* system is involved in induction of competence and transformation and the *blp* system is involved in expression of putative bacteriocins. During initial studies investigating the mechanisms of competence it was not evident that competence and virulence were linked although it had always been tempting to believe that they were. There is now a lot of evidence indicating that the two are intertwined. Components of the *com* system have been identified by STM as being required for bacterial survival in murine models (Hava and Camilli, 2002; Lau *et al.*, 2001). The *lytA* gene, encoding autolysin a known virulence factor in *S. pneumoniae*, has been shown to be part of the *cinA-recA* operon which is induced during competence (Mortier-Barriere *et al.*, 1998). Two microarray studies investigating the transcriptional kinetics during competence development and identifying genes whose expression was dependent on the presence of the signaling peptide (Peterson *et al.*, 2000; Rimini *et al.*, 2000) both identified a subset of genes not previously known to be associated with competence but with possible roles in virulence such as choline-binding proteins. Other TCSTS, such as CiaRH and MicAB are involved in regulation of competence and have a role in virulence.

The *blpHR* system was one of the TCTSTs shown to have a role in virulence in a pneumonia model (Throup *et al.*, 2000). Closer inspection of the neighbouring region in the genome identified a short ORF, *blpC*, encoding a peptide similar to ComC, and two genes, *blpA* and *blpB*, highly related to the ABC transporter ComAB all suggestive of the presence of a second quorum sensing system in *S. pneumoniae*. Microarray-based

expression profiling led to the identification of 16 genes whose expression was highly expressed in a BlpC dependent manner. Their actual function and which are directly involved in virulence is yet to be fully determined. Most encode for relatively small peptides. Seven *blp* ORFs, distributed over three operons, contain a Gly-Gly motif characteristic of peptides that are secreted and processed by ComA/B-like proteins and sequence similarity searches give putative bacteriocin-like function to many of these (de Saizieu *et al.*, 2000). Bacteriocins are compounds that are produced by bacteria that selectively inhibit or kill closely related bacteria or species. One can see a role for these bacteriocins is aiding strain colonization of the nasopharynx where there is competition with other pneumococcal strains and other bacterial species but as the lung parenchyma is usually sterile what role they play in lung infection is not apparent. The authors also found strain-dependent sequence variation in the Blp genes and demonstrated that there was no *blp* regulon cross-induction between strains possessing different *blp* pherotypes. Bacteriocin diversity has been proposed to be the outcome of intense microbial competition and emergence of resistance in target organisms. The related Com system also displays sequence variation within the signal peptide, ComC, and in its proposed target, the N-terminal of the histidine kinase, ComD (Havarstein *et al.*, 1997). Such strain-specific peptide pheromones may allow *S. pneumoniae* clones to co-ordinate their responses to external stimuli even in the presence of other *S. pneumoniae* strains or other related streptococcal species.

The MicAB two-component system provides another example of this link between competence and virulence. Oxygen limitation in a microaerobic atmosphere abolishes development of competence. Studies of oxygen-independent mutant strains demonstrated the involvement of the CiaRH and ComDE TCSTS in this regulation but

the oxygen-sensing mechanism was not apparent (Echenique *et al.*, 2000). PAS domains perceive cell energetic status by sensing oxygen, redox potential, ligands, proton motive force and light (Taylor and Zhulin, 1999). Redox sensing and signal transduction via TCSTS carrying PAS domains is one strategy used by bacteria for adaptation to variations in oxygen concentrations. Genome analysis found the histidine kinase MicB to be the only PAS domain containing protein in *S. pneumoniae* (TIGR4 genome) (Echenique and Trombe, 2001a). Mutation of the PAS domain in MicB abolishes the kinase activity and allows expression of competence under microaerophilic conditions indicating that at least one function of the MicAB system is the perception of signals produced in response to oxygen deprivation and their transduction, culminating in the repression of competence. Whether this repression of the *comCDE* operon is via control of CiaRH regulation or a complementary pathway has yet to be established. The response regulator MicA has been shown to be essential for bacterial growth as insertion mutations into the *micA* gene could not be obtained in any genetic background (Lange *et al.*, 1999; Throup *et al.*, 2000). The MicAB system more recently been shown to be involved in virulence in a murine model of pneumonia and it has been proposed that it is involved in the adaptive response of the bacteria to changes in oxygen levels during the course of infection (Kadioglu *et al.*, 2003).

The functions of the PnpRS and VncRS TCSTS are less clear but have been investigated. Based on similarity to *B. subtilis* PhoP/R and the fact that the genes encoding it are upstream of genes encoding a phosphate specific transporter (*pstABC*), the PnpRS system was initially thought to be regulated by/or linked to phosphate uptake. However the fact that the expression of the *pst* genes did not require an intact PnpRS system and mutations in *pnpRS* did not result in a phenotype similar to the *pst*

mutants suggests that there is no functional relationship between the two loci (Novak *et al.*, 1999a). The finding that *pnpRS* expression was not responsive to phosphate levels also supports this hypothesis (Orihuela *et al.*, 2001). A more recent study demonstrates that, at least in serotype 4, this TCSTS has a role in virulence, regulation of the genes *psaA, psaB* and *psaC* and resistance to oxidative stress. There was no evidence in this study, however, that the PnpRS was involved in virulence or regulated the same set of genes in two other serotypes (2 and 3) (McCluskey *et al.*, 2004).

The VncRS story is more complicated. This TCSTS was originally believed to be part of a third quorum sensing-like system that included an upstream gene encoding the effector signal peptide, Pep27, and three genes upstream to this ORF encoding a putative ABC transporter (VexABC). The whole system was proposed to be associated with vancomycin tolerance and programmed bacterial cell death (Novak *et al.*, 1999a; Novak *et al.*, 2000). A more recent study though, disputes these findings. The authors of this study failed to demonstrate tolerance to vancomycin in either *vncS* or *pep27* mutants. They also demonstrated by microarray transcriptional analysis that inactivation of *vncS* or vnc*R* failed to alter expression of genes known to contribute to autolysis that had been implicated in earlier work. A role for VncRS in transcriptional regulation of the *vex* gene cluster was found though the function of the products encoded by the *vex* genes has yet to be confirmed (Robertson *et al.*, 2002).

1.5.3 Phase variation

Phase variation is another mechanism of gene regulation employed by *S. pneumoniae*. It is a strategy used by many bacterial species that can provide a large repertoire of phenotypes from within a clonal population and potentially provides adaptation to

changing environmental conditions (Synder *et al.*, 2001). The molecular mechanisms behind phase variation have been best worked out in *Neisseria spp.* and are associated with reversible changes within simple DNA repeats, composed of repeated sequence motifs of less than 10 nucleotides. These repeats are located either within open reading frames (ORFs) or within promoters. Alterations in the number of repeats located within ORFs alter the relative translational reading frame of the sequence 5' and 3' of the repeat (Stern *et al.*, 1986; Stibitz *et al.*, 1989). Changes in the length of repeats located within promoters alter the relative position of promoter components and thereby influence transcription (Sarkari *et al.*, 1994; van der Ende *et al.*, 1995).

A total of 397 (18%) of genes in the TIGR4 genome have short sequence repeats (Tettelin *et al.*, 2001) and are potential candidates for undergoing phase variation in a similar way. Of these 25 are directly related to virulence. The two best described phase variants in *S. pneumoniae* are characterised by differences in colony opacity appearance. Transparent variants have less polysaccharide capsule, more cell wall teichoic acid, more choline-binding protein A (CbpA) and less pneumococcal surface protein A (pspA) than their opaque counterparts (Kim and Weiser, 1998; Weiser and Kapoor, 1999). The genes from the teichoic acid and capsule pathways that are associated with colony opacity variation are amongst those identified in the TIGR4 genome to iterative DNA motifs. These phenotypic variants are also associated with differences in virulence. Only the opaque variant of several strains was able to cause sepsis in a bacteraemia model (Kim and Weiser, 1998) whereas transparent variants were more efficient at adhering to human endothelial and epithelial cell lines and colonising the nasopharynx in animal models (Cundell *et al.*, 1995c; Weiser *et al.*, 1996). Transparent variants are selected for during carriage in an infant rat model

(Weiser *et al.*, 1994) compared to a selection for opaque variants during otitis media and in septicaemia (Tong *et al.*, 2001). Despite knowing that *S. pneumoniae* undergoes phase variation the actual genetic mechanisms that contribute to it in *S. pneumoniae* have not been demonstrated.

1.5.4 Inter strain variation in virulence gene regulation

Hand in hand with the increase in our knowledge of virulence gene regulation and regulatory networks in *S. pneumoniae* over the last three years has been the growing realization that virulence determinants and even regulatory pathways can differ from strain to strain. Pneumococcal pathogenesis work to date has utilized a variety of different strain backgrounds, e.g 2, 3, 4, 6, 19F and 22. There has been little consistency in gene naming or annotation. Even since the availability of the TIGR4 genome there has been a lack in attempt to annotate genes according to TIGR4 locus names. This produces some difficulty with cross-referencing work performed in different strain backgrounds. Yet despite these hurdles it is becoming apparent that inter-serotype/strain differences do occur for isolated virulence gene distribution and resultant product function. Some virulence factors are not conserved across serotype. There are at least 10 genes identified as important for virulence of serotype 4 in a murine model of pneumonia that are not present in D39 or R6 (Hava and Camilli, 2002; Tettelin *et al.*, 2001), e.g. the *rlrA* locus and *pia* locus (this work and (Brown *et al.*, 2001)). There are an increasing number of examples of factors that are involved in disease production in one genetic background but not others. These include regulatory components, which may also imply that there are different targets for these regulators in different strains. Inactivation of MicB resulted in an attenuated phenotype with respect to causing pneumonia in a serotype 6B or 2 background (Kadioglu *et al.*, 2003) but had a wild-

type phenotype in a serotype 3 background (Throup *et al.*, 2000). Similarly mutations in the response regulator *SP0661* resulted in different phenotypes with respect to ability to cause disease in different murine models in four serotypes, 2, 3, 4 and 22 (this work) (Blue and Mitchell, 2003; Lange *et al.*, 1999). Inactivation of *regR* resulted in an attenuated phenotype in a 6B background but not in D39 (Chapuy-Regaud *et al.*, 2003) despite the fact that at least one of its targets, *hyl*, is present in both strains. This data and the finding that there is diversity in signaling pheromones in *S. pneumoniae* suggests that each *S. pneumoniae* clone is potentially unique in its exact quota of virulence determinants and co-ordination of their expression.

AIMS OF THE PROJECT

At the time this project was initiated there was very little published about regulation of virulence gene expression in *S. pneumoniae* and nothing was known about virulence gene regulation *in vivo*. My ultimate goal was to investigate the regulation and expression of one or more pneumococcal virulence genes *in vitro* and *in vivo*. Despite the apparent lack of data on virulence gene regulation at the time, there were several simultaneous advances in the field in general that aided development of this project. The *S. pneumoniae* genome sequence had just been completed (Tettelin *et al.*, 2001). One STM screen had been published (Polissi *et al.*, 1998), another followed shortly afterwards (Lau *et al.*, 2001) and a third was completed by a colleague whilst I was working in the Camilli laboratory (Hava and Camilli, 2002). There was an increasing amount being elucidated on the quorum sensing and regulatory system controlling induction of competence and transformation in *S. pneumoniae*, that gave insights into some regulatory mechanisms employed by this pathogen. Finally there had been some exciting applications of new *in vivo* techniques (RIVET) and microarray technology

that allowed investigation of the temporal expression of virulence determinants *in vivo* in other pathogens (Lee *et al.*, 1999; Lee *et al.*, 2001) and identification of one regulon within *S. pneumoniae* (de Saizieu *et al.*, 2000).

The main aim of the project was to investigate virulence gene regulation in *S. pneumoniae*. To achieve this the following objectives were set:

1. To assess the feasibility of RIVET for use in *S. pneumoniae* to study the expression of virulence factors, initially *in vitro*, and then *in vivo*. Pneumolysin was chosen as the first virulence factor to study.
2. To use different models of infection to determine the requirement of virulence gene regulators identified by STM.
3. To identify genes under the control of the regulators of interest.
4. To perform functional analysis of the virulence genes controlled by the subset of regulators analysed in (2) and (3).
5. To examine the distribution of regulators and the genes which they control in clinical isolates.

CHAPTER 2 MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Bacterial strains and plasmids

Strains, plasmids, and primers used in this study are listed in Tables 2.1, 2.2, 2.3 and 2.4. The parental *S. pneumoniae* strains used in genetic manipulations in this study are either derivatives of TIGR4, a serotype 4 clinical isolate, of AC365, a serotype 9V *S. pneumoniae* clinical isolate or of D39. AC1365 was a gift from Dr J Weiser, University of Pennsylvania Medical School, Philadelphia, Pennsylvania, USA. D39, serotype 2 *S. pneumoniae* strain, was a gift from Dr J Brown, University College Hospital, London, UK. The collection of clinical strains for detection of the *rlrA* islet was a gift from Dr A. Brueggemann, University of Oxford, Oxford, UK.

2.1.2 Media and culture conditions

E.coli strains were grown in Luria-Bertani (LB) broth or on LB agar at 37°C. *S. pneumoniae* strains were grown at 37°C and 5-6% CO_2 on Columbia agar supplemented with 5% horse blood or in Todd-Hewitt broth plus 5% yeast extract (THY), and supplemented with Oxyrase (5 μg ml^{-1}) for serotype 4 *S. pneumoniae* and 0.8% maltose when indicated. Where indicated *lacZ* reporter *S. pneumoniae* strains were grown in THY broth without Oxyrase in a) ambient air (20% oxygen, 0.03% carbon dioxide), b) 10% carbon dioxide, 10% hydrogen and 80% nitrogen (generated by an anaerobic chamber), c) 5% carbon dioxide, 20% oxygen or d) 10% carbon dioxide, 20% oxygen (generated in a micro-processor controlled incubator). Unless otherwise stated, antibiotic concentrations used in this study were as follows:

A) Agar plates: chloramphenicol (Cm) 4 μg ml^{-1}, streptomycin (Sm) 100 μg ml^{-1}, and spectinomycin (Spc) 200 μg ml^{-1} for *S. pneumoniae*; Amp 50-100 μg ml^{-1} , Cm 10 μg ml^{-1} and Spc 100 μg ml^{-1} for *E. coli.*

B) Liquid cultures: chloramphenicol (Cm) 2 μg ml^{-1}, streptomycin (Sm) 50 μg ml^{-1}, and spectinomycin (Spc) 100 μg ml^{-1} for *S. pneumoniae*; Amp 50 μg ml^{-1} , Cm 5 μg ml^{-1} and Spc 50 μg ml^{-1} for *E. coli.*

2.1.3 Reagents and antibodies

Unless otherwise stated, chemicals were purchased from Sigma and enzymes from Gibco-BRL or New England Biolabs and were used with buffers according to the manufacturers instructions.

Rabbit anti-resolvase antibody was a gift from Dr A Camilli (Tufts University, Boston, USA) and was used at dilutions 1:5000. Rhodamine-conjugated phalloidin (Molecular probes) was used at a dilution of 1:50. Cyanine 2 (Cy2)–conjugated donkey anti-rabbit antibodies (Jackson Immunoresearch Laboratories) were used at a dilution of 1:200. Anti-capsule serotype 2 and 4 were used at dilutions of 1:500 - 1:2000.

2.1.4 Eucaryotic cells and culture conditions

A549 cells, a human lung epithelial cell line were grown according to ATCC guidelines in Dulbecco's modified Eagle medium (DMEM; Gibco-BRL) supplemented with 10% foetal bovine serum (FBS) and 2mM glutamine at 37°C and 5% CO_2. These were seeded into 24-well tissue culture plates and allowed to adhere for 48h prior to the binding assay.

2.2 *IN VIVO* STUDIES

2.2.1 Mouse infections and competition experiments

In all animal infections 6 to 10 week old female Swiss Webster (USA) or CD1 (UK) mice were used (Taconic or Charles River Laboratories in either USA or UK respectively). Prior to competition experiments *magellan2* insertion mutations or other mutations were backcrossed into AC353 as described (Hava and Camilli, 2002). Bacteria were prepared for inoculation as follows. Mutant or wild-type strains were grown during the day in THY broth supplemented with the appropriate antibiotic until an OD_{600} was reached. Cultures were then adjusted to an appropriate concentration, split into aliquots, immediately frozen and stored with 20% glycerol at -80°C. The number of colony forming units per millilitre (cfu/ml) was confirmed by serial dilution. Further adjustment to an appropriate input dose was made at the time of inoculation of the mice. In all experiments mice were lightly anaesthetized with methoxyflurane or halothane prior to inoculation and were sacrificed by CO_2 asphyxiation at the end of the infection.

For lung infection 40 µl of the bacterial suspension was inoculated intranasally (approximately $1x10^7$ CFU). The infection was allowed to proceed for 44 hours, at which time mice were euthanased, lungs aseptically removed, mechanically homogenized and serial dilutions plated on blood agar plates. Approximately $5 x10^5$ bacteria were inoculated for *i.p.* infections and bacteria were recovered from the bloodstream by cardiac puncture after 20 hours. To assess nasopharyngeal carriage 10 µl of a bacterial suspension was inoculated intra-nasally (approximately $1x10^8$ CFU) and bacteria were recovered at 7 days by washing the nasopharynx with 400 µl of sterile phosphate-buffered saline.

For single infections the input dose was calculated by serial dilution of the inoculum and the absolute output number, measured as CFU/gram of tissue, was calculated from serial dilutions of the homogenate and knowledge of the weight of tissue dissected. For resolution assays the percentage of Spc^s CFU was determined in the input inoculum and output homogenate by plating recovered bacteria on blood agar plates supplemented with Sm and subsequently replica plating colonies to either Sm or Sm Spc blood agar plates. For competition experiments mutant and wild-type bacteria were mixed in a 1:1 ratio and inoculated at doses described. The input ratio of mutant:wild-type bacteria inoculated into the mice was determined as described below. In parallel to all infections, the same mutant:wild-type mixture was inoculated into 10mls of THY supplemented with appropriate antibiotic and Oxyrase (5 $\mu g\ ml^{-1}$) and grown to mid/late exponential phase to assess *in vitro* growth.

For each mouse competition experiment and *in vitro* competition growth experiments, the ratio of mutant to wild-type bacteria recovered was determined by plating recovered bacteria on blood agar plates supplemented with Sm and subsequently replica plating colonies to either Sm and Sm Cm, or Sm and Sm Spc blood agar plates. *In vivo* competition indices (CI) were calculated as the ratio of mutant to wild-type bacteria recovered from each animal adjusted by the input ratio.

2.3 MANIPULATION OF DNA AND RNA

DNA was electrophoresed through 0.6-1.6% Tris-acetate EDTA buffer (TAE; 40mM tris-acetate, 1mM EDTA, pH 8.0) gels. Gels were stained with ethidium bromide and DNA was visualized in a UV transilluminator. DNA fragment size and concentration was determined by comparison with Hyperladder (Bioline). Isolation of DNA

fragments, ligation into plasmid vectors, bacterial transformation and isolation of recombinant clones were carried out according to standard protocols (Sambrook *et al.*, 1998).

2.3.1 Preparation of genomic and plasmid DNA

Genomic DNA was prepared from bacteria grown overnight on blood agar plates using the DNAEasy Tissue kit (Qiagen) according to the manufacturers tissue preparation protocol. Plasmid DNA was prepared using the QIAprep Spin Miniprep kit and QIAfilter Plasmid Midi kit (Qiagen) according to the manufacturers instructions.

2.3.2 Polymerase chain reaction (PCR)

Primers used in this study are listed in Tables 2.1 and 2.2. Unless otherwise noted, all PCR reactions were performed in reaction buffer containing 1x *Taq* reaction buffer (Promega), 250μM dNTPs, 1μM of each primer, and a 10:1 mix of *Taq* and *Pfu* DNA polymerases. Reaction conditions consisted of 25 cycles of 95°C – 30s, 50 to 52°C – 30s, and 72°C – 30s/kb of DNA, followed by a 10 min post-dwell amplification at 72°C.

2.3.3 Nucleotide sequencing and sequence analysis

Nucleotide sequencing was carried out by the Physiology Department, Tufts University, Boston, USA or by MGW-Biotech (Eberrberg, Germany). Nucleotide sequences were analyzed and translated using DNA Strider. Nucleotide and amino acid sequence alignments and database searches were performed using the BLAST server at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/blast/).

2.3.4 *In vitro* transposon mutagenesis

In vitro magellan2 transposition reactions were carried out with purified MarC9 transposase, 500 ng of target AC353 genomic DNA and 1 μg of each of pEMcat or pEMspc derivative separately, essentially as described (Lampe *et al.*, 1996). Reactions were ethanol precipitated and resuspended in gap repair buffer [50mM Tris (pH 7.8), 10mM $MgCl_2$, 1mM DTT, 100nM dNTPs, and 50 ng of BSA]. Repair of transposition product gaps was performed as described (Akerley *et al.*, 1998), except that *E. coli* DNA ligase (NEB) was used in place of T4 DNA ligase. Repaired transposition products were transformed into naturally competent AC353 as described (Bricker and Camilli, 1999).

2.3.5 Plasmid and bacterial strain construction

(1) Construction of pCH48 and CH49, the *S. pneumoniae* strain harbouring the *res1-aad9-res1* cassette integrated into the chromosome

pCH48 is a derivative of pAC1000 a suicide vector for *S. pneumoniae* and was constructed by stepwise addition of the components making up the final cassette (Figure 2.1). A *BamHI-SacII* fragment of DNA containing *orf2* (*sp0558*) was excised from pAC637, purified and ligated into *BglII-SacII* digested pAC1000 resulting in pCH3. A *KpnI res1-res1* fragment excised from pAC865 was then ligated into an isolated *KpnI* site within *orf2* in pCH3 to make pCH2. Finally, the Spc^R gene (*aad9)* and its promoter was amplified from pEMSpc with primers SPCF1 and SPCR1, containing terminal *XbaI*, sites and purified. An *XbaI* digested fragment was ligated into the isolated *XbaI* site between the two *res1* sequences of pCH2 resulting in pCH48.

To generate the *S. pneumoniae* strain containing *res1-aad9-res1* stably integrated into the chromosomal copy of *orf2* (*sp0558*), pCH48 was linearised by digestion with *Nco1* and the gel-purified fragment was transformed into competent AC353. The double recombination event was selected for by plating on Spc and confirmed by PCR and DNA sequencing.

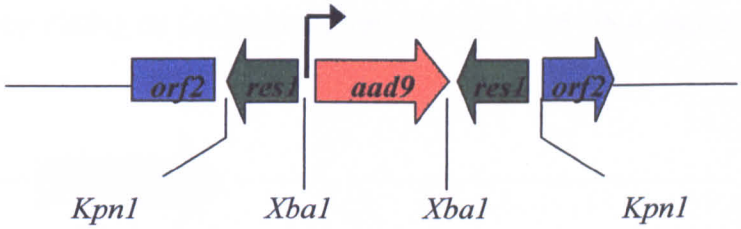


Figure 2.1 *Res1-aad9-res1* antibiotic cassette in pCH48
Stepwise addition of components resulted in the construction of the *res1-aad9-res1* antibiotic cassette in *orf2* (*sp0558*) in pCH48. Restriction enzyme sites are marked. *S. pneumoniae* strains were transformed with linearised plasmid resulting in insertion of the cassette in *orf2* (*sp0558*) onto the chromosome by homologous recombination.

(2) Construction of pCH46 and the *S. pneumoniae* reporter strains CH53 and CH100

To construct a *ply* reporter strain CH53 the $tnpR^{mut135}$ allele was introduced into the CH49 chromosome downstream of *ply*, creating a *ply::* $tnpR^{mut135}$ transcriptional fusion (Figures 2.2 and 3.3). CH49 already contained the *res1-aad9-res1* cassette elsewhere on the chromosome.

DNA fragments containing the 3' end of the *ply* gene and the 685 bp region immediately downstream of the *ply* stop codon were PCR amplified from AC353 with primer pairs 3PLYF1/3PLYR1 and DPLYF1/DPLYR1. The *cat* gene, conferring Cm resistance (Cm^R) on both *E. coli* and *S. pneumoniae*, was PCR amplified from pAC1000 with the primer pairs PCATF1/PCATR1. The $tnpR^{mut135}$ gene was PCR amplified from

pAC676 with primer pairs TNPRF2/TNPRR1. Each of these fragments was subcloned separately into pCR-Script Amp SK (+)(Stratagene) and subsequently inserted into pAC1000 to give pCH46. This construct contained the *ply* sequence and the 5' downstream region to *ply* flanking the coding sequence of $tnpR^{mut135}$ and the *cat* gene. To generate CH53 pCH46 was linearised by digestion with *Xho1* and the gel-purified fragment was transformed into competent CH49. The double recombination event was selected for by plating on Cm and confirmed by PCR and DNA sequencing.

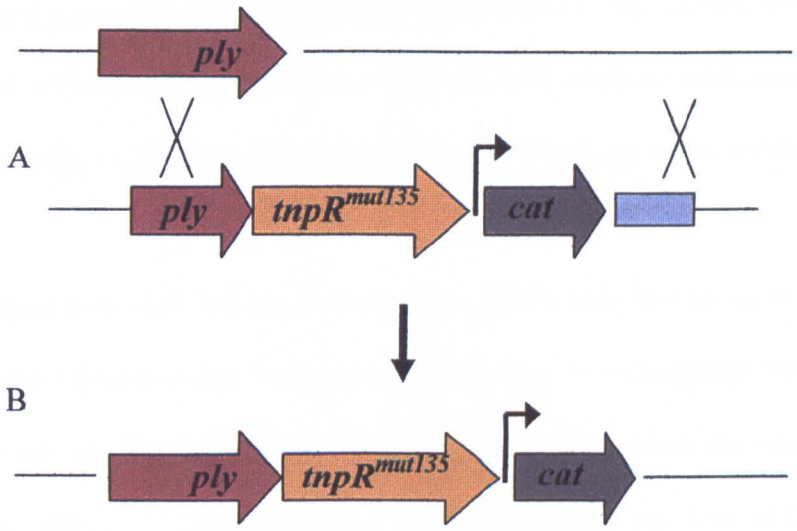


Figure 2.2 *ply::tnpR* transcriptional fusion
The *S. pneumoniae* strain CH53 was constructed by placing *tnpR* adjacent to *ply* on the chromosome.
(A) The plasmid pCH46 contained the last 700 bp of the *ply* sequence and the 689 bp region immediately downstream of the *ply* stop codon (light blue rectangle) flanking the coding sequence of *tnpR* and the *cat* gene.
(B) To generate CH53 pCH46 was linearised and the gel-purified fragment was transformed into competent CH49.

To generate CH100, a *S. pneumoniae* strain with the *tnpR* allele with the *rpoB*-like RBS. Competant CH49 cells were transformed with linearised plasmid in which the $tnpR^{mut135}$ allele had been replaced by the $tnpR^{rpoB}$. $TnpR^{rpoB}$ was constructed by PCR amplifying the gene from AC676 with primer pairs TNPRF3/TNPRR1 and purified.

This time the forward primer contained the sequence for the new *rpoB*-like RBS. A purified *Bgl11-Xba1* digested fragment was ligated into *Bgl11-Xba1* digested pCH46.

(3) Construction of the new $tnpR^M$ gene and $ply::tnpR^M$ fusion strain

The codon usage within the *S. pneumoniae* genome and the *tnpR* gene were compared using the backtranslate program online at http://www.entelechon.com/eng/backtranslattion.html, selecting for the *S. pneumoniae* codon bias table, in conjunction with the total proteome codon usage table from http://wwww.tigr.org/tigr-scrips/CMR2/codontables.spl?project=bsp. Any rare or infrequent codons highlighted by the backtranslate program were changed to a more commonly used codons, selected for by using the TIGR proteome codon usage data.

The resultant new *tnpR* sequence, designated $tnpR^M$, was broken up into seven 100 bp lengths with 18 bp overlaps between each fragment. Each fragment would serve as the sequence for an oligo (ACTM1-7) from which to reconstruct the $tnpR^M$ gene. A tail containing *Bgl11* or *Xba1* sites was incorporated into the first (ACTM1) and last (ACTM7) oligos respectively. The native *tnpR* RBS was replaced by an *rpoB*-like RBS sequence in ACTM1. Further forward (ACTMF) and reverse (ACTMR) primers to amplify the entire fragment were also designed. The 100 bp oligos were synthesised by the primer and sequencing service of the Department of Physiology, Tufts University, Boston, USA.

The oligos were annealed by heating the reaction mixture [1x *pfu* reaction buffer (Promega) and1μM of ACTM1-7, ACTMF and ACTMR] to 98°C and slow cooling to 50°C. The gaps were filled in by extension at 65°C for 5 minutes after addition of

250μM dNTPs and *Pfu* DNA polymerase. The products were purified using Qia-quick PCR purification Kit (QIAGEN), eluting with 0.1X elution buffer. Repair of the product gaps was performed as described (Akerley *et al.*, 1998), except that a thermo-ligase, *Taq* DNA ligase (NEB), was used in place of T4 DNA ligase and ligation was performed for 1 hour at 65^{o}C. After further product purification the resultant putative *tnpR^{M}* gene was amplified with primer set ACTMF/ACTMR and cloned into pCRScript-SK+ vector and transformed into *E.coli* DH5α. Multiple clones were sequenced and one with the correct *tnpR^{M}*, CH95, was selected.

A *Bgl11-Xba1* digested fragment, containing *tnpR^{M}*, was excised from pCH95 and ligated into *Bgl11-Xba1* digested pCH46 to create pCH103. This construct contained the 3'*ply* sequence and the 5' downstream region to *ply* flanking the coding sequence of *tnpR^{M}* and the *cat* gene. To generate the reporter strain CH104, an *S. pneumoniae* strain habouring the *tnpR^{M}* downstream of *ply*, pCH103 was linearised by digestion with *Xho1* and the gel-purified fragment was transformed into competent CH49. The double recombination event was selected for by plating on Cm and confirmed by PCR and DNA sequencing.

(4) Construction of maltose inducible/over-expressing *S. pneumoniae* strains

To construct strains that express *lacZ, tnpR^{M}, mgrA, rlrA* or *sp0661* from an inducible promoter, the coding sequence of each gene was introduced into *S. pneumoniae* maltose locus downstream of *malM* (Acebo *et al.*, 2000b; Puyet and Espinosa, 1993) (Figure 2.3). DNA fragments containing the 3' end of the *malM* gene and the 5' end of the *malP* gene were PCR amplified from AC353 with primer pairs MALMFX/MALMRP and MALPF2/MALPRP. The *cat* gene, was PCR amplified from pAC1000 with the

primer pairs PCATF1/PCATR1. The *tnpR*^{M} gene, cloned into pCR-Script Amp SK (+)(Stratagene), was constructed as previously described. The coding sequence of *lacZ* was amplified from pEVP3 with primer sets LACFPN/LACZR1 and LACFRB/LACZR1. The coding sequence of *mgrA, rlrA and sp0661* genes were PCR amplified from AC353 with the primer sets SP1800F/SP1800R, RLRAFR/RLRARX and 0661F/0661R. In the cases of *mgrA, rlrA, sp0661* and one of the *lacZ* alleles the Shine-Dalgarno sequence of the *S. pneumoniae rpoB* gene was engineered into the forward primer sequence to allow a high level of translation efficiency of each respective gene at the maltose locus. Each of these fragments was sub cloned separately into pCR-Script Amp SK (+)(Stratagene). All DNA fragments were subsequently inserted into pAC1000 to give pAC1472, pCH73, pCH75, pCH101 (Table 2.4). The final constructs contained the 3'*malM* sequence and the 5' *malP* sequence flanking the coding sequence of each gene under study and the *cat* gene.

To generate the *S. pneumoniae* strains over-expressing the chosen genes each of the plasmids above were linearised by digestion with *Xho1* and the gel-purified fragment was transformed into naturally competent AC353. The double recombination event was selected for by plating on Cm and confirmed by PCR and DNA sequencing. To generate the serotype 2 *S. pneumoniae* strain over-expressing *mgrA* the serotype 4 *mgrA* gene sequence was excised from pAC1472 and replaced with an equivalent *mgrA* gene sequence amplified from D39. The resultant plasmid was linearised by digestion with *Xho1* and the gel-purified fragment was transformed into naturally competent D39.

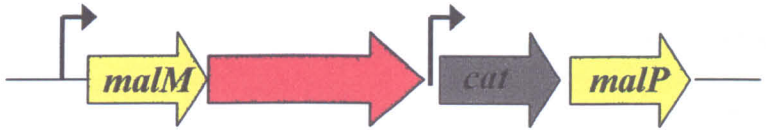


Figure 2.3 Maltose inducible/over-expressing *S. pneumoniae* strains
Maltose inducible *S. pneumoniae* strains were constructed by placing the required gene (marked red; *lacZ, tnpR*^{M}, *mgrA, rlrA* or *sp0661*) downstream of *malM* as a transcriptional fusion on the chromosome. Plasmids were constructed with the coding sequence of each gene under study flanked by DNA fragments containing the last 500 bp of *malM* and the *cat* gene and the first 500 bp of *malP*. To generate the *S. pneumoniae* strains over-expressing the chosen genes each of these plasmids were linearised and the gel-purified fragment was transformed into naturally competent AC353.

(5) Construction of *sp0156* and *sp0661* deletion strains

Constructs necessary to introduce unmarked in-frame deletions in *sp0156* and *sp0661* into the *S. pneumoniae* chromosome were generated by splice overlap extension using a PCR fragment corresponding to the nucleotide sequence upstream region to the coding sequence and a PCR fragment corresponding to the nucleotide sequence downstream of the stop codon. In each case, the upstream fragment was generated by PCR from AC353 using the primers pairs 156SF1/156SR1 or 0661SF1/0661SR1. The downstream PCR products were amplified using 156SF2/156SR3 or 0661SF2/0661SR2. Following amplification, each PCR product was purified using the StrataPrep PCR clean-up kit (Stratagene).

Next, the upstream and downstream PCR products were fused in a second PCR reaction. This reaction contained 5 µl of each PCR product, F1 and R2 primers (above), 250 nM dNTPs, and 1.5 µl of a 10:1 mixture of *Taq* and *Pfu* DNA polymerases in 1X *Taq* buffer, and consisted of 13 cycles of 95°C – 30s, 45°C – 30s, and 72°C – 1 min. The resulting PCR products were gel purified as described above and ligated into pCR-

Script Amp SK (+)(Stratagene). The cloned fragment was subsequently liberated by digestion with *BamH1* and *SacII*, and cloned into pAC1000, digested in a similar manner, to generate pCH117 and pCH130. The final plasmids were introduced into AC353 by transformation and integrated plasmids were selected by plating on Cm and confirmed by PCR.

To generate strains that harboured the appropriate deletion on the chromosome, the *S. pneumoniae* strains containing the singly integrated plasmid were grown in THY in the absence of selective pressure for the plasmid. After growth, cultures were diluted and plated on blood agar plates containing Sm, and resulting colonies were replica-plated to blood agar plates containing Cm. Strains that failed to form colonies on Cm plates were identified and examined for the presence of the deletion by PCR and ultimately confirmed by DNA sequencing. In this manner, CH119, containing Δ*sp0661* was isolated. Repeated efforts were made to obtain the Δ*sp0156* strain, however all strains that lost the integrated plasmid regenerated the wild-type copy of *sp0156*. Therefore, *magellan2* transposon insertion strain, CH108, was used in competition assays.

(6) Construction of *mgrA* deletion strains

AC1500 was constructed by replacing 1299 bp of the coding sequence of *mgrA* (corresponding to bp 180 to stop codon) with the Spc^R gene (*aad9*) and its promoter amplified from pEMSpc (Figure 2.4). DNA fragments containing the region 5' to *mgrA* gene and ending 180 bp into the coding sequence of the gene and the region 3' to the *mgrA* gene were PCR amplified from AC353 using primer pairs 180AF1/1800R8 and DN1800/180SRX. The *aad9* gene and its promoter, conferring spectinomycin resistance to *E. coli* and *S. pneumoniae*, was PCR amplified from pEMSpc using primer

pairs SPCFN/SPCRS. Each of these fragments was subcloned separately into pCR-Script Amp SK(+)(Stratagene) and subsequently inserted into pAC1000 to give the allelic exchange vector pAC1499. To generate AC1500, pAC1499 was linearised by digestion with *Xho*I and the gel-purified fragment was transformed into naturally competent AC353 as previously described. The double recombination event was selected for by plating on Spc and confirmed by PCR and DNA sequencing.

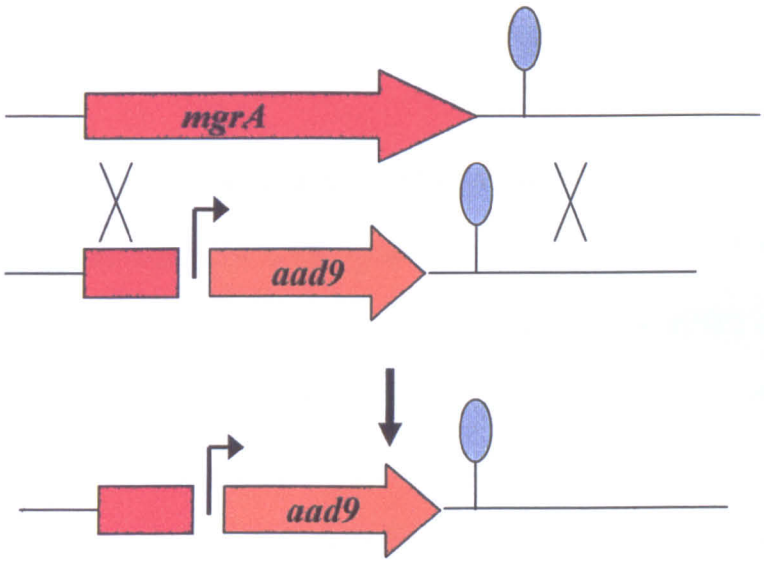


Figure 2.4 *MgrA* deletion strain
The *S. pneumoniae* strain AC1500 was constructed by 1299 bp of the coding sequence of *mgrA* with *aad9* ($Spec^R$ gene). Terminator is marked as a blue hoop.
(A) The plasmid pAC1000 contained the first 180 bp of *mgrA* and the region immediately downstream of *mgrA* flanking *aad9* and its promoter.
(B) To generate AC1000 pAC1000 was linearised and the gel-purified fragment was transformed into competent AC353.

To construct serotype 2 strains with a deletion-insertion of a Spc^r gene in place of *mgrA* PCR amplification of the mutated region in the serotype 4 strain, AC1500, was performed using primer pairs 180AF1/180SRX and the purified amplicon transformed into *S. pneumoniae* D39 or other serotype 2 background. The double recombination event was selected for by plating on Spc and confirmed by PCR and DNA sequencing.

(7) Construction of CH209, Δ*mgrA* Δ*rlrA* double mutant

To construct CH207, the TIGR4 strain with a deletion-insertion of a Spc^r gene in place of *mgrA* and a *magellan2* transposon insertion in *rlrA*, PCR amplification of the mutated *mgrA* region in AC1500 was performed using primer pairs 180AF1/180SRX and the purified amplicon transformed into competent *rlrA::magellan2* cells, AC1213. The double recombination event was selected for by plating on Spc and confirmed by PCR and DNA sequencing.

(8) Construction of *lacZ* transcriptional fusion strains

Plasmids containing the *lacZ* gene transcriptionally fused to either *rlrA* or *srtD* were derivatives of pAC1000. The *lacZ* gene was PCR amplified from pEVP3 using primer pairs LACFRB/LACZR1, *srtD* and *rlrA* from AC353 with SRTDF1/SRTDR1 and RLFF1/RLFR1 respectively and *erm*, a gene conferring erythromycin resistance to *S. pneumoniae*, from pCH6 with primer pairs ERMF3/ERMR3. The primers had terminal restriction sites incorporated into each one to allow sequential addition of genes to the pAC1000. Each purified DNA fragment was initially inserted into pCR2.1-TOPO, using the TA cloning kit (Invitrogen) according to the manufacturer's instructions. From there they were excised with appropriate restriction enzymes and ligated into the pAC1000 backbone. Two final plasmid constructs were made, one with *rlrA* fused to *lacZ* and a second with *srtD* fused to *lacZ* upstream of the *erm* gene. These plasmids were introduced into competent AC353 by transformation to generate CH210, the *S. pneumoniae* strain with a plasmid insertion resulting a *srtD::lacZ* at the chromosomal *srtD* site and CH214, the *S. pneumoniae* strain with a plasmid insertion resulting a *rlrA::lacZ* at the chromosomal *rlrA* site. Pneumococcal strains CH211, CH212, CH213

and CH215 are strains with either a *srtD::lacZ* fusion or *rlrA::lacZ* with deletions in one or both *mgrA* and *rlrA* (table 2.2). These were created by PCR amplification of the mutated *mgrA* region in AC1500 or the mutated *rlrA* region in AC1213 with primer pairs 180AF1/180SRX and RLRN1/RLRN2 and the purified amplicon was transformed into competent CH210 or CH214. The double recombination event was selected for by plating on Spc for *mgrA* insertion-deletion or Cm for *rlrA* transposon mutation and confirmed by PCR.

(9) Construction of non-serotype 4 *rlrA* or *rrgA* mutant strains

To generate serotype 2 or 9V *S. pneumoniae* strains with mutations in either *rlrA* or *rrgA* PCR amplification of the mutated region in the mutant serotype 4 strain was performed using primer pairs RLRAFR/RLRARX or RRGAF1/RRGAR1 and the purified amplicon transformed into required serotype 2 or 9V background. The double recombination event was selected for by plating on Cm for *rlrA* or Spc for *rrgA* confirmed by PCR.

(10) Construction of serotype 2 strains possessing *rlrA* islet

Competent D39 cells were transformed with genomic DNA from CH155, a serotype 4 *S. pneumoniae* strain with a magellan5 transposon insertion in one of the IS1167 elements flanking the *rlrA* islet. The double recombination event was selected for by plating on Spc and confirmed by PCR.

2.3.6 *rlrA* islet detection in clinical isolates

Clinical isolates were provided by Dr A. Brueggemann, University of Oxford, Oxford, UK. Chromosomal DNA was purified from each strain using the DNAEasy kit as

previously described. Dilutions of each sample were made and spotted onto Hybond N+ nitrocellulose membranes (Amersham Pharmacia) and UV crosslinked. Membranes were processed for hybridisation and developed using Alk Phos direct labeling and detection kit (Amersham Pharmacia) according to the manufacturers recommendations. Hybridisation was carried out at 52°C after which membranes were washed twice in primary buffer for 10 minutes at 52°C and twice in secondary buffer for 5 minutes at room temperature. Nucleotide probes were prepared by PCR from AC353 genomic DNA and primer sets RLRAF2/T7, SRTDF/SRTDR and RPOBPF/RPOBPR.

2.3.7 Preparation of RNA

Total RNA was isolated from 5 ml aliquots of *S. pneumoniae* grown in THY broth supplemented with Oxyrase (5 µl ml^{-1}) with the addition of 0.8% maltose where indicated. Cells were pelleted at 4 °C and immediately frozen. Bacterial pellets were thawed on ice and treated with 400 mg ml^{-1} lysozyme in 100 µl of PBS for 5 min at room temperature. RNA was isolated with Qiagen Rneasy kit in accordance with the manufacturers (Qiagen) recommendations. Samples were treated on the column with DNase I (Qiagen), as recommended by the manufacturer.

2.3.8 Microarray Probe synthesis, hybridisations and data Analysis.

Probe synthesis and Hybridisations. RNA was converted to cDNA in 20 µl reactions by combining 1 µg of RNA and 0.5 µg of random hexamers (Amersham), heating to 65°C for 10 minutes, and then snap cooling the reactions on ice. The following was then added: 2 µl of 0.1 M dithiothreitol (DTT), 0.5 µl of 10 mM dNTPs, 4 µl of 5x RT buffer (Invitrogen) and 1 µl (200U) of Superscript II (Gibco BRL). This mixture was incubated at 42°C for 150 minutes. RNA was hydrolyzed with 1 µl of 1 M NaOH at

65°C for 10 minutes and neutralized with 1 μl of 1 M HCl. Samples were purified over a Qia-quick PCR column (Qiagen) according to the manufacturer's instructions and eluted with 40 μl elution buffer. Amino-allyl dUTP was incorporated into the cDNA samples as follows. For each sample, 40 μl of the eluted DNA was incubated for 5 minutes at 99°C and then for 5 minutes on ice. 5 μl of 10x random octamer buffer (NEB 1550-2), 3 μl of dNTP/dUTP mix (3 mM dGTP, dATP, dCTP; 1.8 mM aa-dUTP (Sigma-Aldrich A-0410), 1.2 mM dTTP] and 2 μl of Exo$^-$ Klenow (NEB) were added, and the mixture was incubated for 150 minutes at 37°C and then stored at 4°C overnight. Free amines were removed with a Qia-quick PCR purification kit (Qiagen), and the eluted sample was dried in a speed-vac. Samples were resuspended in 4.5 μl of dH_2O and incubated with 1μM of either Cy3 or Cy5 monofunctional reactive dye (Amersham) for 1 h at room temperature in the dark. The time point samples were incubated with Cy5, and the reference samples were incubated with Cy3. Reference samples were prepared by reverse transcribing a pool of RNA comprised of equimolar aliquots of RNA isolated from wild-type (AC353) bacteria at each OD. The reactions were quenched with 4.5 μl of 4 M hydroxylamine for 15 minutes at room temperature, and then each Cy5-labeled sample was mixed with a Cy3-labeled reference. Unincorporated dye was removed with a Qia-quick PCR purification kit, and probes were eluted with 40μl of EB and were dried in a speed-vac. To hybridize, the samples were resuspended in 11.3 μl of TE, pH 7.5, 1 μl of 10 mg ml^{-1} 1 yeast tRNA, 2.25 μl of 20x SSC and 0.45 μl of 10% SDS. The mixture was heated to 99°C for 2 minutes and immediately centrifuged for 2 minutes at maximum speed. The probe was applied to a microarray, which is described elsewhere http://falkow.stanford.edu/whatwedo/supplementarydata/pub7/MicroarrayDesign.pdf and incubated for at least 24 hours at 60°C.

Data Analysis. Arrays were scanned using a GenePix 4000A scanner (Axon Instruments, Union City, CA) and images were analyzed with GenePix Pro 3.0 software. Microarray data were stored in the Stanford Microarray Database (Gollub *et al.*, 2003) and are publicly available (http://genome-www.stanford.edu/microarray). The data were filtered to remove poor quality measurements (for example, spots affected by scratches on the array were not considered) and the red::green ratios were log2 transformed. Genes for which reliable measurements were obtained for over 80% of the arrays in the resulting dataset were selected and organized by hierarchical clustering using the CLUSTER program and were viewed in TREEVIEW (Eisen *et al.*, 1998). Statistical analysis was performed using Significance Analysis of Microarrays (SAM) (Tusher *et al.*, 2001).

2.3.9 Ribonuclease protection assays

Template DNA for the generation of riboprobes was PCR amplified with the primer sets below (Table 2.1). The resulting products were purified with a QIAquick PCR purification kit, subsequently cloned into pGEM-T (Promega) and confirmed by PCR with both a T7 or SP6 primer and a primer specific for the cloned insert. These plasmids were used as templates for generation of riboprobes (Merrell and Camilli, 2000). Synthesized probes were purified on a 4% denaturing polyacrylamide gel containing 7M urea. Ribonuclease protection assays (RPAs) were carried out as described by the manufacturer with an RPA11 kit (Ambion) in triplicate with at least 3 independently isolated RNA samples. The protected fragments were visualized by exposing each gel to a phosphor-imaging screen (Kodak) and analyzed with a Storm 860 scanner and IQMac V 1.2 imaging software. The relative amount of each protected

fragment in each assay was normalized to the amount of *rpoB* protected RNA in each lane.

Table 2.1 Primer pairs and plasmids used for synthesis of RPA probes

Probe	Primer set	pGEM-T Plasmid
0663	663PF/663PR	pCH144
0668	668PF/668PR	pCH145
1800	1800PF2/1800PR	PCH55
1803	1803PF/1803PR	pCH147
1804	1804PF/1804PR	pCH148
lytA	LYTAF1/LYTAR1	pCH30
ply	PLYPF1/PLYPR1	pCH31
rlrA	RLRAF2/ RLRAR7	pAC1279
rpoB	RPOBF3/RPOBR2	pAC1286
rrgA	RRGAF3/RRGAR3	pAC1280
rrgB	RRGBF2/RRGBR1	pAC1281
rrgC	RRGCF2/RRGCR2	pAC1282
srtB	SRTBF2/SRTBR1	pAC1283
srtC	SRTCF2/SRTCR2	pAC1284
strD	SRTDF2/SRTDR2	pAC1285

2.4 BIOCHEMICAL STUDIES

2.4.1 Resolution assays

For *in vitro* assays *S. pneumoniae* strains were grown at 37°C and 5-6% CO_2 in THY broth supplemented with Oxyrase (5 μg ml^{-1}) and Cm 4 μg ml^{-1} to maintain the presence of the *tnpR* allele. Aliquots of this culture were serially diluted and ~ 200 CFU were plated onto blood agar plates and incubated overnight. The resulting colonies were replica plated onto blood agar plus Spc 200 μg ml^{-1} and plain blood agar to determine

the percentage of colonies that were Spc^S. *In vivo* assays were performed as described previously.

2.4.2 SDS-PAGE and Western blotting

Proteins were separated by electrophoresis using BioRad Ready Gels in SDS-PAGE running buffer. When indicated, gels were stained with Coomassie brilliant blue and dried. Western blots were performed using the ECL Plus Western Blotting Analysis System as described by the manufacturer (Amersham Pharmacia). Proteins were transferred to Hybond ECL nitrocellulose membranes in transfer buffer, blocked for 1 h in PBS-T (1X PBS, 0.1% Tween-20) plus 5% milk, and incubated overnight at 4°C with anti-resolvase antibody (1:2000) in PBS-T plus 5% milk. After washing, membranes were incubated with horseradish peroxidase linked anti-rabbit antibody diluted 1:2000 in PBS-T for 1 h at RT. Membranes were subsequently processed for detection as described in the manufacturer's instructions.

2.4.3 β-galactosidase assays

S. pneumoniae cells were grown to mid-exponential growth (OD_{600}= 0.3 to 0.4) in THY broth, THY 0.8% maltose or THY sucrose supplemented with appropriate antibiotics then chilled to ~ 4°C. 1 ml of cells were collected by centrifugation and the pellets immediately frozen and stored at –80 °C. Thawed pellets were resuspended to one-tenth of the original volume in 0.1 M sodium phosphate buffer, pH 7.5. Lysis was accomplished by adding Triton-X100 to 0.1% and incubating for 10 minutes at 37 °C. β-galactosidase activity was determined as described by Miller (Miller, 1972) in assay reactions containing 1 ml of reaction buffer [1 mM $MgCl_2$, 50 mM mercaptoethanol, 0.1 M sodium phosphate buffer, pH 7.5, and 0.8 mg ml^{-1} ONPG] and a variable volume

(1/20-1/5 volume) of the sample. Enzyme activity was followed at 420 nm at room temperature (~ 25 °C) and was calculated according to Miller and reported in Miller units (Miller, 1972).

2.5 INFECTION OF CULTURED CELLS

2.5.1 A549 adherence assays

A549 cells were seeded into 24-well tissue culture plates 48 h prior to the binding assays and grown to ~90-95% confluence. *S. pneumoniae* cells were grown to mid-exponential growth (OD_{600}= 0.3 to 0.4), washed once with PBS, adjusted to an OD_{600} = 0.3 in PBS and diluted 1:10 in Ham's F12K tissue culture medium plus 10% foetal bovine serum. Aliquots of 350 µl were added to the monolayers of A549 cells at an MOI of ~10:1. Bacteria were incubated with A549 cells for 30-40 minutes at 37°C in 5% CO_2, at which time the culture fluid was removed from each well and the monolayers were washed 3 times with PBS (pH 7.4) to remove non-adherent bacteria. For enumeration of adherent and/or internalized bacteria, epithelial cells were detached from the wells by treatment with 200 µl of 0.25% trypsin-1mM EDTA and lysed by the addition of 800 µl of ice-cold 0.025% Triton X-100. Appropriate dilutions were plated on blood agar plates containing Sm to count the number of bacteria adherent to the eukaryotic cells. The titer of adherent bacteria for each strain was compared to the input titer and the percentage of adherent bacteria was determined.

For microscopy A549 monolayers were grown on coverslips in 24-well tissue culture plates. Infected cell layers on coverslips were fixed and prepared as described below after the 30-40 minute incubation and washing with PBS. All experiments were

performed in quadruplicate and each experiment was replicated 3 times on different days.

2.6 MICROSCOPY

2.6.1 Sample preparation and fluorescent microscopy

For immunofluorescence, cell monolayers on coverslips were fixed in 3% PFA in PBS, pH 7.4, for 15 minutes at room temperature and washed three times in PBS. The remaining free aldehyde groups in PFA were quenched with 10 mM NH_4Cl_2 before labeling. Labeling antibodies were diluted in 10% horse serum, 0.1% saponin in PBS. Coverslips were washed twice in PBS containing 0.1% saponin, incubated for 30 minutes in the dark at room temperature with primary antibodies, washed twice in PBS containing 0.1% saponin and then incubated for 30 minutes in the dark at room temperature with secondary antibodies. Finally, coverslips were washed twice with 0.1% saponin in PBS, once in PBS and once in H_2O and mounted on Mowiol.

Samples were analysed by epifluorescence microscopy (BX50; Olympus Optical Co.). Images were captured using a cooled 12-bit high-resolution monochrome digital camera and analysis imaging software (Soft Imaging System). Images were processed and false coloured where appropriate using Adobe Photoshop software (Adobe Systems).

Table 2.2 Primers used in this study

Primer Name	Sequence (5' to 3')
0661F	CGCGGATCCAAAGGAGAATCATCATGACCTACACAATCTTA ATC
0661SF1	ATTACCATCAAGACTATCTC
0661SF2	CAGAGCATGTGATTAAGAAAGAGTTGGG
0661SR1	TCTTAATCACATGCTCTGCTCCTTAC
0661SR2	GCCCCAACTGGAGTTGAT
0661R	CCCTCTAGATCATCGGTCTTCTCCCTTCTT
663PF	GTCAGGTAGTGGAAGATGT
663PR	ATGAAGAGTAGGATGGACA
668PF	GGTGAAAACGAGAAATGTA
668PR	GGCTCATGGAAATCACTGT
156SF1	CAAGTCATTGCTCAAGGC
156SF2	GAGAAAATGGCCCAGCAGCTCTTGCTTTC
156SR1	CTGCTGGGCCATTTTCTCCCTTTCTACT
156SR3	ATATGTACCTCCATGATTTTATCTATAATC
1800PF2	CATGCTATACCTATTCTTTGT
1800PR	TAATACGACTCACTATAGGTACGATAGAGCGAAGATGAAC
1800R8	GCATGCCATGGTTAAGAAGAATGAAAAATCAAG
180AF1	GTTTGGATCCTTTAAAATTAAAAATATAATCTT
180SRX	GGGACTCGAGATGAAACACAAGGAATGGCA
1803PF	ATGGAATGGCTTAAACA
1803PR	ATCCAGCTGCAACTCCC
1804PF	GTAGCTCACGAAATCAAAG
1804PR	AATATAAAGCTGGAACATT
3PLYF1	GCCAAGTCTATCTCAAGTTG
3PLYR1	GGGCCAGATCTCTAGTCATTTTCTACCTTAT
ACTM1	GCGTAGATGTTTACAAAGGAGAATAATTATGCGCTTATTCGG TTATGCACGTGTTAGTACCTCTCAACAAAGTCTTGACATCCA GGTCAGAGCTTTGA
ACTM2	AGCAAATCGAGACCTTTTCGGTCACTACTAGATCCAGAAGCC TTATCAGTAAAAATTCGATTAGCCTTCACACCTGCATCTTTC AAAGCTCTGACCTGGA
ACTM3	AAAAGGTCTCGATTTGCTACGAATGAAAGTGGAAGAAGGTG ATGTTATTCTAGTGAAAAAGTTGGATFCGCTTGGGTCGTGAT ACCGCTGATATGATCCAG
ACTM4	TTTTACCCATTTCTCCGTCTGTACTAATTCCATCATCAATAAA TCTAATAGATACTCCTTGTGCATCAAATTCTTTAATAAGCTG GATCATATCAGCGGT
ACTM5	ACGGAGAAATGGGTAAAATGGTTGTTACTATTTTATCTGCCG TTGCCCAAGCTGAAAGACAACGCATTCTTGAGCGCACTAAT GAGGGAAGACAGGAGGC

ACTM6	GCCGAGGCCTTGTTGCCACATATTCAAGACAGCGTCGCGGTC AATCTTTCGTTTACGACCAAAGACTACACCTTTTGCCATAGC CTCCTGTCTTCCCTCA
ACTM7	TGGCAACAAGGCCTCGGCGCTTCACATATTTCAAAGACAAT GAACATCGCTAGATCAACAGTTTATAAAGTAATCAATGAAT CAAATTAATCTAGACCGC
ACTMF	GCGTAGATGTTTACAAAGGAG
ACTMR	GCGGTCTAGATTAATTTGATT
DN1800	GGATGCATGCCAAAAATAACAAAAAAAC
DPLYF1	CCCGGCTGCAGTAGCATGCGAGAGGAGAATGCTTGCGAC
DPLYR1	CGCGACTCGAGTAGAAAGTTTCAGCCAAGTT
ERMF3	TGCCTCTAGAGGGAACAGAGATTTTGAC
ERMR3	GAGGCATGCTTATTTCCTCCCGTTAAA
LACFPN	CGCGGATCCAAAGGTGGTGAACTACTG
LACFRB	CGCGGATCCAAAGGAGAATCATATGGAAGTTACTGACGTAA GA
LACZR1	GGGTCTAGATTATTTTTGACACCAGACCA
LYTAF1	CGCACACTCAACTGGGAATC
LYTAR1	TGTTTGGTTGGTTATTCGTG
MALFX	CCCTCGAGTGAAAGCTATCGTGAGCAATT
MALPF2	CCCTCTAGAGAGCATGCGACAATAATCAGGAGACAAC
MALPRP	CCGCGGCTCGAGTTCAAGAGGCCATTTTTCAAG
MALRP	CCGAGCTCAAGATCTGGATCCTTATTTCTTTAAATCTACC
OUTPFL	ACTACATCACTAACGCCATTG
OUTXYL	AATTCTAGAAGGCATACCAGT
PCATF1	CCCGGTCTAGAGTCGACGGTATCGATAAGCT
PCATR1	CCGGCGCATGCTTATAAAAGCCAGTCATTAG
PLYPF1	GGCAAATAAAGCAGTAAATG
PLYPR1	ACAAGGTCTCATCCACTACG
RLFF1	ATACCGGCGGAACAGCACTGACTTTCTTA
RLFR1	CGCGGATCCTTATAACAAATAGTGAGCC
RLRAF2	TTACATGCTGTTTTATCAATAA
RLRAFR	CGCGGATCCAAAGGAGAATCATCATGCTAAACAAATACATT GA
RLRAR7	AGTAGAAAGAAGCGGAGTATT
RLRARX	CCCTCTAGATTATAACAAATAGTGAGCCTT
RPOBF3	TGCTTATGACTTGGCAGCAG
RPOBR3	GGCTTTCAATGCTTTCAATC
RRGAF1	AGTCTTAACAGGGAGACACAC
RRGAF3	CACTTTTATACGCTTTTGCTA
RRGAR1	GCAACATTCCGATACCACCAG
RRGAR3	TAATACGACTCACTATAGGTGCCATCCGTATTGTTTTTC
RRGBF2	AAACTATCATTGAAAGGGGAG
RRGBR1	TAATACGACTCACTATAGGGGCATTGCCCTGAGAGTTTA
RRGCF2	GGCTGCGATTATGGGTATT

RRGCR2	TAATACGACTCACTATAGGGGTCATCTCAAACGAAGTCT
SP1800F	CGCGGATCCAAAGGAGAATCATCATGAGAGATTTATTATCT AAAAAAAG
SP1800R	CCCTCTAGATTACTCATCTAATCGAATAAA
SPCF1	CCGCTCTAGAACTAGTGGATCC
SPCR1	CCCTCTAGACAATTTTTTTATAATTTTTTTAATCTG
SRTBF2	AGGACTGGGATTCTGATTTA
SRTBR1	TAATACGACTCACTATAGGATCGCCACTCACTACATTATT
SRTCF2	GATTCTTTTATGGATTATTCG
SRTCR2	TAATACGACTCACTATAGGGACGCCTTTCTTTTTCTCTTG
SRTDF1	TTTCCGCGGGGTTTATTTGGGAGCAG
SRTDF2	GCGGTCATCCTTCTCTTGCT
SRTDR1	CGCGGATCCTTATTTCCCTCGTAGTAAA
SRTDR2	TAATACGACTCACTATAGGGTCGTCAGACACTTGGTAAT
TNPRF2	GGCCCAGATCTTTGAGATACATTTTTATGC
TNPRR1	GGCCCGCATGCTCTCTAGATTAGTTGCTTTCATTTATTA
XYL1	AGCATACTCCAACTCATAAAT
XYL2	TCAGGAAAATCTGACTTTAACA

Table 2.3. Strains used in this study

Strains	Relevant Genotype or Phenotype	Source or Reference
E. coli strains		
DH5αλ	F- D(*lacZYA-argF*) U169 *recA1 endA1 hsdR1 supE44 thi-1 gyrA96 relA1*	(Hanahan, 1983; Kolter *et al.*, 1978)
XL-1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*[F'*proAB lacI*^{q}*ZDM15 Tn10 (Tet*^{r}*)*]	Stratagene
CH48	DH5α contains pCH48	Chapter 3
AC1287	DH5α contains pAC1287	Chapter 4
AC1288	DH5α, contains pAC1288	Chapter 4
AC1289	DH5α, contains pAC1289	Chapter 4
AC1290	DH5α, contains pAC1290	Chapter 4
AC1291	DH5α, contains pAC1291	Chapter 4
AC1292	XL-1 Blue, contains pAC1292	Chapter 4
V. cholerae strains		
AC66	C6709-1 (El Tor) *lacZ::res-tet-res*. Sm^R Tc^R	(Camilli and Mekalanos, 1995)
AC311	C6709 lacZ::res-tet-res vieS'::pIVET5. Sm^R Ap^R Tc^S, LacZ+. Resolved strain.	Andrew Camilli
S. pneumoniae strains		
TIGR4	Wild-type Type 4 encapsulated strain	Ingeborg Aaberge
AC353	Transformed Sm^R derivative of TIGR4	(Bricker and Camilli)
CH49	AC353 *orf2::res1-aad9-res1*. Sm^R Spc^R	Chapter 3
CH53	AC353 *ply::tnpR*^{mut135}, *orf2::res1-aad9-res1*. Sm^R Spc^R Cm^R	Chapter 3
CH79	AC353 *malM::lacZ::cat::malP*. Sm^R Cm^R	Chapter 3
CH80	AC353 *malM::lacZ*^{mutRBS}*::cat::malP*. Sm^R Cm^R	Chapter 3
CH100	AC353 *ply::tnpR*^{mutrpoB}, *orf2::res1-aad9-res1*. Sm^R Spc^R Cm^R	Chapter 3
CH102	AC353 *malM::tnpR*^{M}*::cat::malP. orf2::res1-aad9-res1* Sm^R Spc^R Cm^R	Chapter 3
CH104	AC353 *ply::tnpR*^{M}, *orf2::res1-aad9-res1*. Sm^R Spc^R Cm^R	Chapter 3
CH107	AC353 *sp0247::magellan2*. Sm^R Cm^R	Chapter 4
CH108	AC353 *sp0156::magellan2*. Sm^R Cm^R	Chapter 4
CH119	AC353 Δ*sp0661*. Sm^R	Chapter 4
CH155	AC353 *IS1167:: magellan5*. Spc^R	Dave Hava
CH184	AC1365 *rrgA::magellan5*. Spc^R	Chapter 6
CH185	AC1365 *rlrA::magellan2*. Cm^R	Chapter 6

CH207	D39 Δ*mgrA*. Sm^R Spc^R	Chapter 5
CH208	D39 Δ*mgrA*. Spc^R	Chapter 5
CH209	AC353 Δ*mgrA* Δ*rlrA*. Sm^R Cm^R Spc^R	Chapter 5
CH210	AC353 *srtD::lacZ*. Sm^R Cm^R Ery^R	Chapter 5
CH211	AC353 *srtD::lacZ* Δ*mgrA*. Sm^R Spc^R Cm^R Ery^R	Chapter 5
CH212	AC353 *srtD::lacZ* Δ*rlrA*. Sm^R Cm^R Ery^R	Chapter 5
CH213	AC353 *srtD::lacZ* Δ*rlrA* Δ*mgrA*. Sm^R Spc^R Cm^R Ery^R	Chapter 5
CH214	AC353 *rlrA::lacZ*. Sm^R Cm^R	Chapter 5
CH215	AC353 *rlrA::lacZ* Δ*mgrA*. Sm^R Cm^R Spc^R	Chapter 5
CH216	D39 *malM::mgrA::cat::malP*. Sm^R Cm^R	Chapter 6
CH226	D39 *rlrA* islet *IS1167:: magellan5*. Spc^R Sm^R	Chapter 6
CH232	D39 *rlrA* islet *IS1167:: magellan5 rlrA::magellan2*. Spc^R Cm^R Sm^R	Chapter 6
CH244	D39 *rlrA* islet *IS1167:: magellan5* Δ*mgrA*. Sm^R Cm^R Spc^R	Chapter 6
CH246	D39 *rlrA* islet *rrgA::magellan5*. Spc^R Sm^R	Chapter 6
CH254	D39 *rlrA* islet *IS1167:: magellan5 malM::mgrA::cat::malP*. Sm^R Cm^R Spc^R	Chapter 6
AC1213	AC353 *rlrA::magellan2*. Cm^R	(Hava *et al.*, 2003a)
AC1214	AC353 *srtD::magellan2*. Cm^R	Chapter 6
AC1215	AC353 *rrgA::magellan5*. Spc^R	Chapter 6
AC1272	AC353 *mgrA::magellan2*. Sm^R Cm^R	Chapter 4
AC1278	AC353 *malM::rlrA::cat::malP*. Sm^R Cm^R	(Hava *et al.*, 2003a)
AC1365	P12 Type 9V serotype strain	Jeff Weiser
AC1481	AC353 *malM::mgrA::cat::malP*. Sm^R Cm^R	Chapter 4
AC1500	AC353 Δ*mgrA*. Sm^R Spc^R	Chapter 4
STM90	AC353 *sp0661::magellan2*. Sm^R Cm^R	(Hava and Camilli, 2002)
STM206	AC353 *mgrA 5 UTR::magellan2*. Sm^R Cm^R	(Hava and Camilli, 2002)

Table 2.4 Plasmids used in this study

Plasmid	Relevant Genotype or Phenotype	Source or Reference
Plasmids		
pEMCat	Contains *magellan2*. Ap^R Cm^R	(Akerley *et al.*, 1998)
pEMSpc	Contains *magellan5*. Ap^R Spc^R	(Martin *et al.*, 2000)
pGEM-T	Cloning vector. Ap^R	Promega
pCRScript	Cloning vector. Ap^R	Stratagene
pEVP3	*lacZ* transcriptional fusion suicide vector. Cm^R	(Claverys *et al.*)
pAC46	pACYC184 Δ*tet::* containing $tnpR^{mut135}$. Cm^R	Andrew Camilli
pAC637	pCRSript::*sp0558*. Ap^R	Andrew Camilli
pAC676	pIVET5::*irgA'::tnpR*^{mut135} . Ap^R	(Lee *et al.*, 1999)
pAC865	pCRSript-SK+::*res1-res1*. Ap^R	Andrew Camilli
pAC1000	pEVP3 Δ*lacZ*; suicide vector. Cm^R	(Hava *et al.*, 2003a)
pAC1279	pGEM-T *rlrA* RPA probe. Ap^r	Chapter 4
pAC1280	pGEM-T *rrgA* RPA probe. Ap^r	Chapter 4
pAC1281	pGEM-T *rrgB* RPA probe. Ap^r	Chapter 4
pAC1282	pGEM-T *rrgC* RPA probe. Ap^r	Chapter 4
pAC1283	pGEM-T *srtB* RPA probe. Ap^r	Chapter 4
pAC1284	pGEM-T *srtC* RPA probe. Ap^r	Chapter 4
pAC1285	pGEM-T *srtD* RPA probe. Ap^r	Chapter 4
pAC1286	pGEM-T *rpoB* RPA probe, Ap^R	(Hava *et al.*, 2003a)
pAC1472	pAC1000 '*malM-mgrA-cat-malP'*. Sm^R Cm^R	Chapter 2
pAC1499	pAC1000 Δ*mgrA'::aad9'*. Cm^R Spc^R	Chapter 2
pCH2	pAC1000 *orf2::res1-res1*. Cm^R	Chapter 2
pCH3	pAC1000 *orf2*. Cm^R	Chapter 2
pCH30	pGEM-T *lytA* RPA probe. Ap^R	Chapter 2
pCH31	pGEM-T *ply* RPA probe. Ap^R	Chapter 2
pCH46	pAC1000 '*ply-tnpR*^{mut135}*-cat-3'ply*. Cm^R	Chapter 3
pCH48	pAC1000 *orf2::res1-aad9-res1*. Cm^R	Chapter 3
pCH57	pGEM-T *tnpR* RPA probe. Ap^R	Chapter 2
pCH73	pAC1000 '*malM-lacZ-cat-malP'*. Sm^R Cm^R	Chapter 2
pCH75	pAC1000 '*malM-lacZ*^{mutRBS}*-cat-malP'*. Sm^R Cm^R	Chapter 2
pCH84	pAC1000 '*malM::rlrA::cat::malP*'. *Sm*^{R} Cm^R	Chapter 4
pCH95	pCRSript-SK+::*tnpR*^{M}	Chapter 2
pCH101	pAC1000 '*malM-tnpR*^{M}*-cat-malP'* Sm^R Cm^R	Chapter 2

pCH103	pAC1000 '*ply-tnpR*^{M}*-cat-3'ply*. Cm^R	Chapter 2
pCH117	pAC1000 Δ*sp0661*. Cm^R	Chapter 4
pCH130	pAC1000 Δ*sp0156*. Cm^R	Chapter 4
pCH144	pGEM-T *sp0663* RPA probe. Ap^r	Chapter 5
pCH145	pGEM-T *sp0668* RPA probe. Ap^r	Chapter 5
pCH147	pGEM-T *sp1803* RPA probe. Ap^r	Chapter 4
pCH148	pGEM-T *sp1804* RPA probe. Ap^r	Chapter 4

CHAPTER 3 RIVET AND *S. PNEUMONIAE*

3.1 INTRODUCTION

Recombinase-based *in vivo* expression technology or RIVET was originally designed to identify genes that are specifically induced during infection (Camilli *et al.*, 1994). The methodology uses a reporter gene to detect transcription and is based on the loss of an antibiotic resistance, eg. Spectinomycin. The reporter gene, *tnpR*, encodes a site-specific resolvase enzyme derived from Tn*1000*. When its substrate, resolvase recognition sequences flanking an antibiotic resistance gene (*res1-aad9-res1*), is present in the genome, expression of the resolvase leads to permanent excision of this cassette (Figure 3.1). The reaction, referred to as resolution, results in a change in the antibiotic resistance phenotype of the bacteria. The post-resolution antibiotic-sensitive phenotype is stable, inheritable and can be used as a qualitative measure of gene induction during infection.

Although originally developed as a promoter trap in *Vibrio cholerae*, RIVET has been used in other organisms (Lowe *et al.*, 1998; Wang *et al.*, 1996). It has also been adapted in *V. cholerae* to determine the spatiotemporal pattern of transcriptional induction of specific virulence genes (Lee *et al.*, 1998; Lee *et al.*, 1999). RIVET had never been used in *S. pneumoniae* and I aimed to develop the technique for use in this organism to study the expression of virulence factors, initially *in vitro*, and then *in vivo*. Pneumolysin was chosen as the first virulence factor to study.

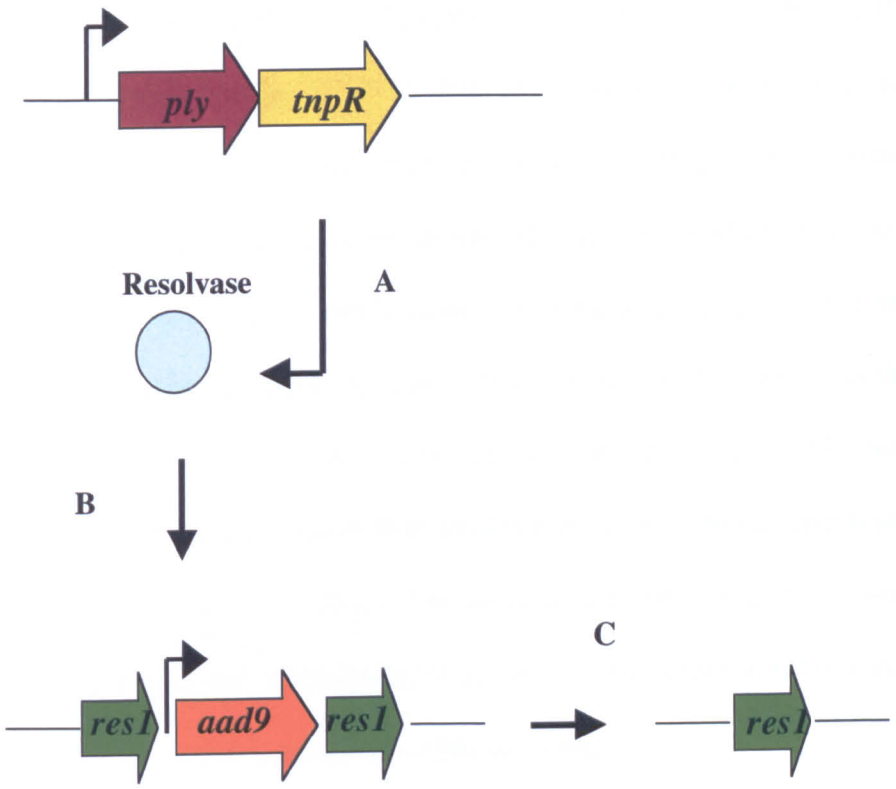


Figure 3.1 Schematic of Recombinase based *in vivo* expression technology (RIVET)

The reporter gene, *tnpR*, encodes a site-specific resolvase enzyme **(A)** derived from Tn*1000*. When its substrate, resolvase recognition sequences flanking an antibiotic resistance gene (*res1-aad9-res1*), is present in the genome expression of the resolvase leads binding to sequences within the *res1* sites **(B)** and permanent excision of this cassette **(C)**. The reaction, referred to as resolution, results in a change in the antibiotic resistance phenotype of the bacteria. The post-resolution antibiotic sensitive phenotype is stable, inheritable and can be used as a qualitative measure of gene induction during infection.

Pneumolysin (Ply) is a thiol-activated multifunctional toxin produced by virtually all clinical isolates of *S. pneumoniae* and is an important pneumococcal virulence factor (Benton *et al.*, 1997; Boulnois *et al.*, 1991a; Canvin *et al.*, 1995; Paton, 1996). It is a cytoplasmic protein (Johnson, 1977) that, at the time of starting this project, was believed to be released on *S. pneumoniae* autolysis, leading to mammalian cell lysis, activation of complement and stimulation of cytokine production (Boulnois *et al.*, 1991b; Mitchell and Andrew, 1997). Autolysis of bacteria by autolysin (LytA), another recognised virulence factor (Berry *et al.*, 1989; Tomasz *et al.*, 1988), was thought to be the likely mechanism by which Ply is released (Ronda *et al.*, 1987). Subsequently, published work has demonstrated that autolysin is not an absolute requirement for Ply release (Balachandran *et al.*, 2001). The toxin is released in significant amounts from some pneumococci during exponential phase growth before autolysis occurs and is equally well released from autolysin-deficient strains.

During growth *in vitro*, cytoplasmic and extracellular Ply is not normally detectable until late log phase (Benton *et al.*, 1997). This observation had been used as evidence that *ply* transcription or release of Ply from the bacterial cell is regulated, but formal published experimental evidence of regulation was lacking. Experiments in animal models of pneumonia suggest that Ply facilitates intrapulmonary bacterial growth and invasion of *S. pneumoniae* into the blood early in disease (Rubins *et al.*, 1995; Rubins *et al.*, 1996). The exact timing of *ply* expression and whether expression occurs in only a subset of all the bacterial cells is not known.

I hypothesised that *ply* (and *lytA*) transcription is activated or at least up-regulated at an early stage of infection. The primary aim of this project was to establish the pattern of

ply expression *in vivo* in murine models of infection with the use of RIVET. Adaptation of the RIVET system for use in studying *ply* expression would provide proof-of-principle for the study of other virulence genes of *S. pneumoniae in vivo*. This chapter describes attempts to adapt RIVET for use in *S. pneumoniae* using *ply* as the reporter gene.

3.2 RESULTS

3.2.1 Determination of pneumolysin gene (*ply*) expression *in vitro*

Preliminary experiments were performed to determine the baseline expression of *ply* and *lytA* in THY broth cultures *in vitro*. Wild-type serotype 4 strain, AC353, was grown in THY broth. To examine the hypothesis that *ply* transcription might be differentially regulated over the course of growth, RNA was harvested from cells at four separate time points on the growth curve. Ribonuclease protection assays (RPAs) were performed probing for either *lytA* or *ply*.

The RPA is a sensitive procedure for the detection and quantification of RNA species (usually mRNA) in a complex sample mixture of total cellular RNA. A labeled radioactive RNA probe is synthesized that is complementary to part of the target RNA to be analyzed. The labeled probe is then mixed with the sample RNA and incubated under conditions that favour hybridization of complementary transcripts. After hybridization, the mixture is treated with ribonuclease to degrade single-stranded, un-hybridised probe. Labelled probe that hybridized to complementary RNA in the sample mixture will be protected from ribonuclease digestion and can be separated on a polyacrylamide gel and visualised by autoradiography. When probe is present in excess over the target fragment in the hybridization reaction, the intensity of the protected

fragment will be directly proportional to the amount of complementary RNA in the sample mixture.

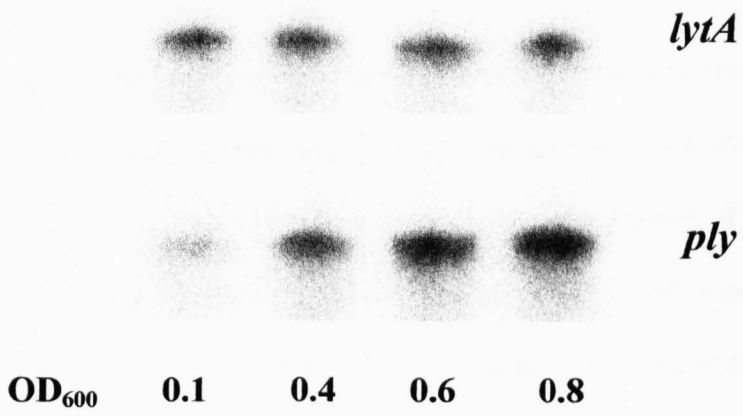


Figure 3.2 Ribonuclease protection assays of *lytA* and *ply* transcripts
RPAs were performed to analyse mRNA levels of the *ply* and *lytA* genes at four different time points during growth, corresponding to OD_{600} 0.1, 0.4, 0.6 and 0.8. Riboprobes to *ply* and *lytA* were generated and hybridised to 5 μg of total *S. pneumoniae* RNA (AC353).

Both the *ply* and *lytA* genes were transcriptionally active. *LytA* had a constant level of expression over the time points tested but there was an increase in *ply* transcript of five fold during exponential growth (Figure 3.2). Although *ply* transcript was detectable even at the earliest time point, it was felt that this would not exclude the use of RIVET as the system could be adapted to decrease the efficiency of translation of the resolvase protein. This is achieved by using a *tnpR* mutant with an altered ribosomal binding (Lee *et al.*, 1999) site in the *ply::tnpR* construct. Use of a mutant allele, with reduced translational efficiency, results in lower levels of resolvase enzyme and hence lower levels of resolution for any given level of transcription. When transcription is low, such as early in growth, resolution is minimal. The increase in transcript levels later in

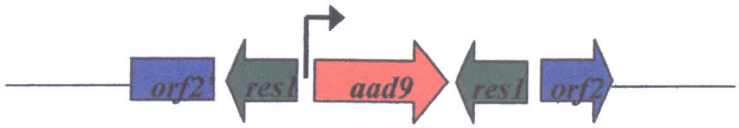
growth should result in increased amounts of enzyme and eventual resolution at these time points. Ideally the system could be tuned to abolish resolution in standard *in vitro* conditions but still allow the detection of any transcriptional induction *in vivo*.

3.2.2 Construction of the *ply* reporter strains and assessment of *in vitro* resolution

Firstly, a pneumococcal strain, CH49, with a resolvable spectinomycin cassette (*res1-aad9-res1*) stably integrated into the chromosome was constructed. The extent of resolution of this strain during *in vitro* growth in the absence of *tnpR* was determined. Cells were grown at 37°C and 5-6% CO_2 in THY broth supplemented with Oxyrase (5 $\mu g\ ml^{-1}$). Aliquots of this culture were taken at time points corresponding to OD_{600} 0.1, 0.4, 0.6 and 0.8. These were serially diluted and ~ 200 CFU were plated onto blood agar plates and incubated overnight. The percentage of colonies that were Spc^S were determined as described in chapter two. This gave a measure of background resolution of the cassette in the absence of the *ply::tnpR* ^{mut13} transcriptional fusion. All colonies were Spc^R indicating that there was no detectable resolution at any time during *in vitro* growth.

Secondly, the resolvase gene, *tnpR*, was inserted immediately downstream of the intact *ply* gene in CH49 creating a transcriptional fusion (Figure 3.3). A *tnpR* allele with reduced translational efficiency, *tnpR*^{mut135}, was used (Lee *et al.*, 1999). The extent of resolution *in vitro* in this *ply:: tnpR* ^{mut13} reporter strain, CH53, was measured during *in vitro* growth, as described above, only cells were grown with the addition of Cm 4 $\mu g\ ml^{-1}$ to maintain the presence of the *tnpR*^{mut135} allele. There was no detectable resolution in this strain at any time point.

Figure 3.3 Construction of RIVET applied to the *ply* gene in *S. pneumoniae*



(A) The *Res1-aad9-res1* antibiotic cassette was cloned in *E. coli*. *S. pneumoniae* strains were transformed with linearised plasmid resulting in insertion of the cassette in *orf2* onto the chromosome by homologous recombination.

(B) The *Ply:: tnpR*^{mut135} fusion was cloned in *E.coli*. *S. pneumoniae* stains were transformed with linearised plasmid resulting in insertion of the fusion onto the chromosome by homologous recombination, giving a stably integrated single copy of *tnpR*^{mut135} immediately downstream of *ply*.

As mentioned above, the *ply* reporter strain above was constructed in two stages. Firstly, AC353 was transformed with linearised DNA from plasmid pCH48, containing the *res1-aad9-res1* cassette. Spectinomycin resistant transformants resulting from double cross-over homologous recombination, were selected. The presence of the cassette was confirmed by PCR and DNA sequencing. Secondly, this strain, CH49, was transformed with linearised plasmid pCH46, which possessed the *ply::tnpR* ^{mut135} fusion and Cm^{R} transformants selected. The insertion of the *ply::tnpR* ^{mut135} fusion into the correct position of the chromosome was confirmed by PCR and sequencing. To assess whether the final strain was immediately resolved upon induction of *ply::tnpR* ^{mut135}, 200 colonies from the second stage transformation were picked and patched onto both blood agar and blood agar supplemented with Spc 100 μg ml^{-1}. All transformants were

Spc^R indicating that none showed immediate resolution on acquiring the resolvase gene. One of these strains, designated CH53, was selected for further study.

3.2.3 Assessment of resolution *in vivo*

The level of gene transcription *in vitro* may not have been adequate to overcome the reduced translational efficiency on the $tnpR^{mut135}$ allele but the level of transcription *in vivo* may be. To determine if this was the case resolution of the *ply* reporter strain, CH53, in a murine model of pneumonia was assessed. Eight mice were inoculated intranasally with ~ 10^7 CFU of CH53. Pairs of mice were euthanased at 2, 8, 24 and 44 hours. The percentage of Spc^R pneumococci recovered at each time point was determined. There was no detectable resolution at any time point.

3.2.4 Investigation of the lack of resolution

There were several possibilities to explain the lack of resolution in the pneumococcal reporter strains including;

1. The antibiotic cassette was unresolvable
2. There was inadequate read-through from the *ply* gene to the *tnpR* gene leading to lack of *tnpR* transcription
3. The RBS mutant was too severe to allow any efficient translation to functional resolvase enzyme
4. The codon usage of *tnpR* was suboptimal for *S. pneumoniae*, resulting in failure to efficiently translate the mRNA
5. Resolvase is non-functional in *S. pneumoniae*

Each of the first four possibilities was looked at in turn.

(1) Cassette function

The antibiotic cassette had been checked by DNA sequencing but it was also confirmed that it was resolvable in a Gram-negative bacterial background. Competent cells were made from CH48, the *E. coli* strain harbouring the complete cassette on a plasmid. These were transformed with pAC46, a plasmid containing $tnpR^{mut135}$ under the control of the pACYC184 *tet* promoter, a strong constitutively active promoter. Transformants selected for by the presence of Cm^R were replica plated to LB Spc 100 μg ml^{-1} to assess for cassette resolution. All transformants were Spc^S. Cassette resolution and plasmid maintenance was confirmed by PCR (Figure 3.4). This showed that the plasmid was maintained and that the cassette had been resolved. This result confirms that the cassette can function normally as a substrate for *tnpR* in an *E. coli* background.

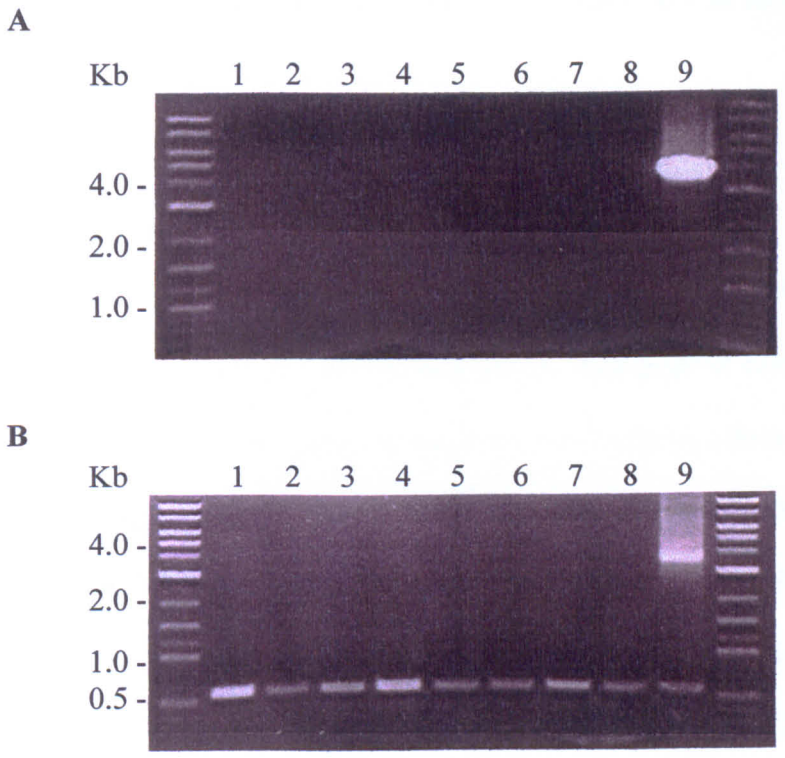


Figure 3.4 PCR to confirm resolution of *res1* cassette in *E. coli*

Figure 3.4 PCR to confirm resolution of *res1* cassette in *E. coli*

(A) Primers ORF1 and SPECF1 were used in a colony PCR to amplify a product that included one *res1* site and the *aad9* gene from 8 test colonies (lanes 1-8) and unresolved pCH48 (lane 9) in order to assess excision of the *aad9* gene.

(B) Primers ORFS1 and ORFS2 were used in a colony PCR to amplify the resolved cassette from 8 test colonies (lanes 1-8) and the unresolved cassette in pCH48 (lane 9) in order to confirm plasmid presence and cassette size.

(2) *TnpR*^{mut135} transcription in *S. pneumoniae*

RPAs were performed to determine the presence of *tnpR*^{mut135} transcript in *S. pneumoniae,* CH53. RNA was harvested from cells grown to mid exponential phase. This was probed for the presence of *tnpR, ply* and *rpoB* transcript. *RpoB* was used as a loading control. Lack of gene transcription did not appear to be the reason for failure of resolution as transcript for each of the three genes were detectable at all time points tested (Figure 3.5).

(3) Translational failure

To determine if *tnpR*^{mut135} transcript was translated into resolvase enzyme, Western blots probing with anti-resolvase antibody were performed on total protein extracted from cells grown to mid log phase. Pure resolvase enzyme, a *V. cholerae* strain possessing the *tnpR* gene as a single copy integrated into the chromosome downstream of a constitutively active promoter (AC311), the parental *V. cholerae* strain habouring the *res-tet-res* cassette alone (AC66) and wild-type *S. pneumoniae* (AC353) were used as controls. The 50μg of protein was loaded in each lane. An identical SDS-PAGE gel was prepared and stained with coomasssie brilliant blue to assess the total protein loading (not shown). There was no resolvase detectable by western blot in CH53, AC66

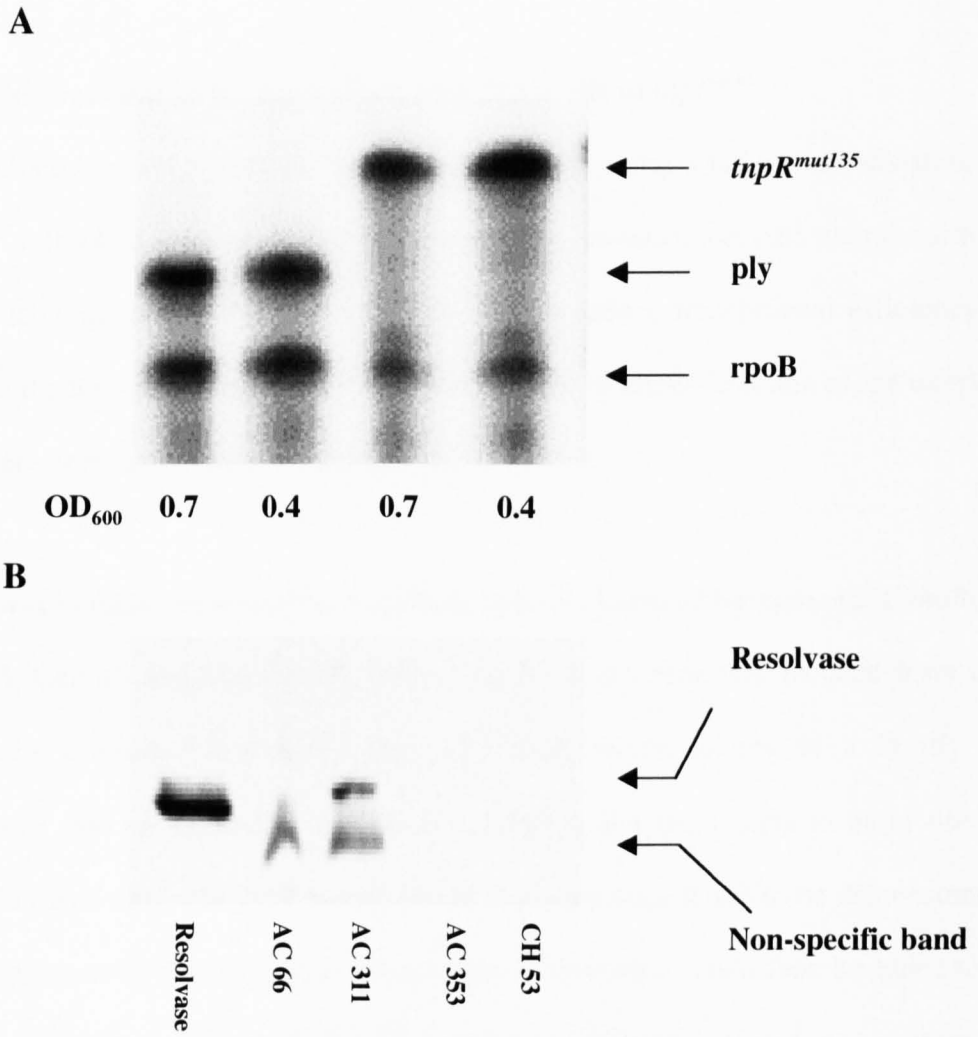


Figure 3.5 RPA for *tnpR*^{mut153} transcript and Western blot for Resolvase protein
(A) RPA to determine the presence of mRNA levels for *ply* and *tnpR*^{mut135} in cells grown to OD_{600} 0.4 and 0.7. Riboprobes to *ply, tnpR* and *rpoB* were generated to hybridise to 5 µg of total *S. pnuemoniae* RNA (AC353). RNA was probed for either *ply* and *rpoB* transcript (lanes 1 and 2) or *tnpR* and *rpoB* transcript (lanes 3 and 4).
A *rpoB* probe was used as a loading control.
(B) Western blot probing total protein extracted from cells grown to mid exponential phase with anti-resolvase antibody; Lane 1, purified resolvase enzyme, lane 2, (AC66) *V. cholerae* strain harbouring the *res-tet-res* cassette alone, lane 3, (AC311) *V. cholerae* strain harbouring *res-tet-res* and *vieS'::tnpR* fusion, lane 4 (AC353) wild-type *S. pneumoniae* and lane 5, (CH53), *S. pneumoniae* strain harbouring *orf2::res1-aad9-res1*and *ply::tnpR*^{mut135}fusion. The 50µg of protein was loaded in each lane. An identical SDS-PAGE gel was prepared and stained with coomasssie brilliant blue to assess the total protein loading (not shown).

or AC353, but detectable resolvase was present in AC311 (Figure 3.5).

3.2.5 Replacement of the Shine-Dalgarno (RBS) site of *tnpR*^{mut135}

The results described above are consistent with there being a failure in translation of *tnpR*^{mut135} mRNA to functional resolvase enzyme in *S. pneumoniae*. Although the aim of using a RBS mutant in the reporter strain was to reduce translational efficiency it appeared the level of translation, if any, was too low to allow detection of the enzyme by Western or measurable resolution *in vitro* on *in vivo*.

RIVET was initially demonstrated in a Gram negative bacterial background (Camilli *et al.*, 1994; Camilli and Mekalanos, 1995). The resolvase gene was isolated from the prokaryotic transposable element Tn*1000* (Tn$\gamma\delta$), which is one of a family of transposons that are related to Tn*3* (Heffron, 1983). For the system to be of use in studying *in vivo* virulence gene expression in *S. pneumoniae* it had to be demonstrated that resolvase could function in *S. pneumoniae*. The system could then be tuned to a level that could then allow use of RIVET to test certain hypotheses in *S. pneumoniae in vivo*. RIVET had been used successfully to identify *in vivo* induced genes in one other Gram positive organism, *Staphylococcus aureus* (Lowe *et al.*, 1998). In this study the Shine-Dalgarno (RBS) site of the *tnpR* had to be replaced with a Gram-positive homologue to improve the translation of the resolvase.

A new *S. pneumoniae* reporter strain, CH100, was constructed in which the *tnpR*^{mut135} RBS and sequence up to the ATG start codon was replaced that found upstream of the *rpoB* gene in this organism. *In vitro* resolution of this strain was assayed as described previously but found to be absent. Western blots were performed on total protein

extracted from cells grown to mid exponential phase. There was no resolvase detectable by western blot in the new strain, CH100 (Figure 3.6).

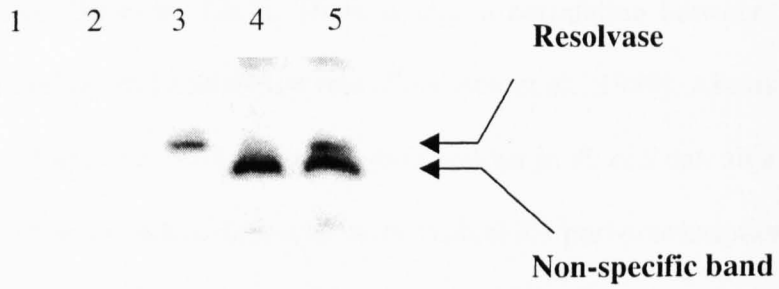


Figure 3.6 Western blot for Resolvase protein in CH100

Western blot probing total protein extracted from cells grown to mid exponential phase with anti-resolvase antibody; Lane 1 (AC353) wild-type *S. pneumoniae*, lane 2 *S. pneumoniae* strain (CH100) harbouring *orf2::res1-aad9-res1* and *ply::tnpR*^{rpoBRBS} fusion. Lane 3 purified resolvase enzyme, lane 4 *V. cholerae* strain (AC66) harbouring the *res-tet-res* cassette alone, lane 5 *V. cholerae* strain (AC311) harbouring *res-tet-res* and *vieS'::tnpR* fusion. 50 µg of total cellular protein was loaded in lanes 1, 2, 4 and 5. An identical SDS-PAGE gel was prepared and stained with coomasssie brilliant blue to assess the total protein loading (not shown).

3.2.6 Assessment of streptococcal codon usage and *tnpR*^{M} gene construction

It appeared that the lack of translation could not be explained by inadequate ribosomal binding to the *tnpR*^{mut135} mRNA. Another possible explanation was that the codon usage in *tnpR*^{mut135} was not optimal for translation in *S. pneumoniae*. Codon usage varies between genes within an organism and from organism to organism (Ikemura, 1981; Sakai *et al.*, 2001). The use of individual codons within one organism can alter the translation rate of the mRNA (Robinson *et al.*, 1984). In *E. coli* synonymous codons are used with varying frequencies in different mRNAs. mRNAs for ribosomal proteins and other highly expressed proteins use a specific subset of codons almost exclusively, so called common codons (Grosjean and Fiers, 1982; Post *et al.*, 1979). The concentrations

of tRNA in *E. coli* have been determined and a correlation has been found between the frequency of codon usage in highly expressed genes and the concentration of the cognate tRNAs (Ikemura, 1981). There is also a correlation between codon usage, tRNA concentration and translation rate (Sorensen *et al.*, 1989). Abolishment of rare codons and replacement with more common codons in *E. coli* can significantly alter mRNA translation to such a degree as to abolish some post-transcriptional regulatory effects (Kuhar *et al.*, 2001).

We hypothesised that the codons present in the native *tnpR* gene were not optimal for mRNA translation in *S. pneumoniae* and that the presence of rare codons for *S. pneumoniae* within the gene was a reason for inadequate translation in the *ply::tnpR* fusion strain. The codon usage within the *S. pneumoniae* genome and the *tnpR* gene were compared by using a combination of the online backtranslation program (http://www.entelechon.com/eng/backtranslattion.html) and the total proteome codon usage table from the TIGR website (http://wwww.tigr.org/tigr-scrips/CMR2/codontables.spl?project=bsp). The frequency of use of any codon for each amino acid in the pneumococcal genome was calculated. A frequency less than 15% was taken to approximate to infrequently used or rarely used codons. 20 out of the 184 codons in *tnpR* fell into this category (Figure 3.7). This number could be sufficient to abolish mRNA translation.

A

```
1    gcc cag atc t tta gga tac att ttt ATG CGA CTT TTT GGT TAC GCA CGG GTA
28   TCA ACC AGC CAG CAA TCT CTC GAT ATT CAG GTT CGG GCA CTC
70   AAA GAC GCA GGC GTG AAA GCA AAT CGC ATC TTT ACT GAC AAG
112  GCA TCG GGC AGT TCA AGC GAT CGG AAA GGG CTG GAC TTG
151  CTG AGG ATG AAG GTG GAG GAA GGT GAC GTC ATC TTG GTG AAG
193  AAA CTT GAC CGC CTT GGG CGC GAT ACT GCT GAC ATG ATC CAG
235  TTA ATA AAA GAG TTT GAC GCC CAA GGT GTA TCC ATT CGG TTT
277  ATT GAT GAC GGA ATC AGT ACC GAT GGG GAG ATG GGT AAA ATG
319  GTT GTC ACT ATT CTA TCT GCA GTG GCC CAG GCA GAA CGA CAG
361  AGA ATA CTA GAG CGT ACC AAT GAA GGT CGC CAA GAG GCA ATG
403  GCA AAA GGA GTT GTT TTT GGT AGA AAA AGA AAA ATA GAT AGA
445  GAT GCA GTA TTA AAT ATG TGG CAA CAG GGG TTA GGT GCC TCA
487  CAT ATA TCA AAA ACA ATG AAT ATT GCT CGT TCA ACA GTA TAT
529  AAA GTA ATA AAT GAA AGC AAC TAA
```

B

Codon	Amino Acid	Triplet frequency in *S. pneumoniae* genome (%)	Triplet frequency for corresponding amino acid (%)
CGG	Arginine	0.2	4.98
AGG	Arginine	0.23	5.62
CTG	Leucine	0.91	8.96
TCG	Serine	0.39	6.05
TCC	Serine	0.52	7.99
ATA	Isoleucine	0.74	10.38
GGG	Glycine	0.84	13.28
GGC	Glycine	0.89	14.04
CGT	Arginine	1.75	42.87

Figure 3.7 Codon usage in *tnpR* and *S. pneumoniae* genome
(A) Nucleotide sequence of *tnpR*. Open reading frame is in upper case. Rare codons are highlighted.
(B) Frequency of codon usage in *S. pneumoniae*. The rare codons selected from the *tnpR* gene and the commonest codon used for Arginine in the whole *S. pneumoniae* genome are reported.

Rather than systematically replace each individual codon in turn it was decided to design and construct a new *tnpR* allele with an altered *rpoB*-like RBS and a *S. pneumoniae*-optimized codon usage. The new allele was constructed as described in chapter two. Seven 100 base oligos (ACTM1-7) that overlapped by 18 bp were designed and synthesized. The first primer included the *rpoB* RBS. Rare codons throughout the gene were changed to ones that were more frequently used in *S. pneumoniae* but which preserved the amino acid sequence of the final protein. The final gene sequence was confirmed by DNA sequencing. A new *ply* reporter strain, CH104, was constructed by replacing the $tnpR^{mut135}$ allele in pCH46 with this new allele ($tnpR^{M}$). Transformation of CH49 with this plasmid resulted in the insertion of the $tnpR^{M}$ allele immediately downstream of *ply* on the chromosome.

3.2.7 Development of a controllable expression system

The work above describes attempts to use RIVET in *S. pneumoniae* using a *ply* reporter system. Whilst performing this work it became apparent that we also needed to develop a way of more directly controlling the expression of *tnpR*, enabling us to titrate the amount of expression to resolution. At the time of this project the only system that had been used to a similar effect in *S. pneumoniae* was the use of the maltose inducible promoter in trans or integration into the maltosaccharide utilization regulon on the chromosome (Acebo *et al.*, 2000a; Nieto *et al.*, 2000). The regulon is organized into three operons (Figure 3.8). Two operons, involved in maltosaccharide uptake (*malXCD*) and catabolism (*malMP*), are divergently transcribed from neighbouring promoters. The third operon (*malAR*) is located downstream and its products are involved in the regulation (Nieto *et al.*, 1997). *MalR* encodes a transcriptional repressor that binds to each promoter. Binding activity is reduced and repression relieved with increasing

concentrations of maltose, whereas addition of sucrose further enhances repression. The operon had previously been manipulated to allow titratable expression of GFP (Acebo *et al.*, 2000b).

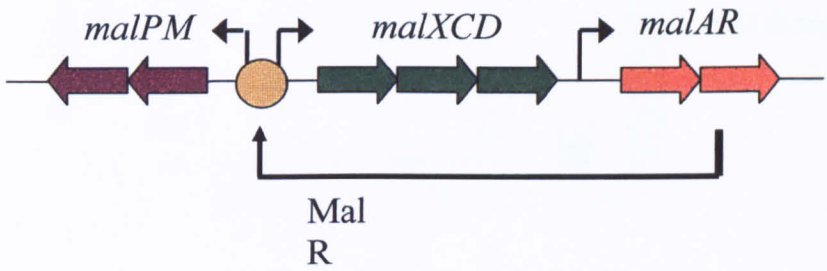


Figure 3.8 The Maltose operon in *S. pneumoniae*
Two operons, involved in maltosaccharide uptake (*malXCD*) and catabolism (*malMP*), are divergently transcribed from neighbouring promoters. The third operon (*malAR*) is located downstream and its products are involved in the regulation. *MalR* encodes for a transcriptional repressor that binds to each promoter. Binding activity is reduced and repression relieved with increasing concentrations of maltose.

To determine if a this regulon could be used to regulate the expression of genes in our hands, pneumococcal strains were constructed with the *lacZ* gene inserted as a single copy between *malM* and *malP* on the pneumococcal chromosome as a pilot study. Two different strains were constructed, CH80 and CH79, one with a *lacZ* gene with the RBS of the *rpoB* gene and the other with a *LacZ* gene with the native RBS from pEVP3, the source plasmid for the *lacZ* gene. AC353, CH79 and CH80 were grown in THY broth alone or THY broth supplemented with sucrose or maltose. Expression of *lacZ* was assayed by measurement of β-galactosidase activity in whole cells and reported in Miller units (Figure 3.9). Measurable β-galactosidase activity in AC353 was negligible. CH79 and CH80 grown in THY or THY sucrose gave increased

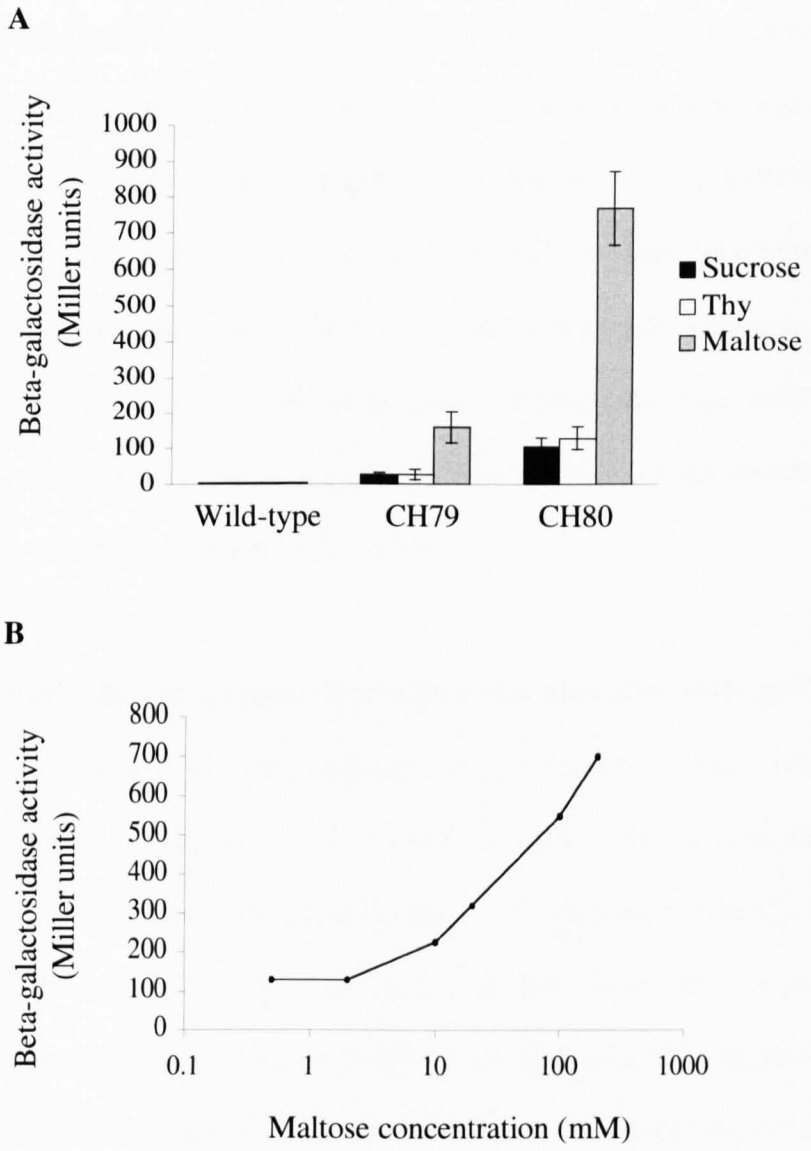


Figure 3.9 β-galactosidase activity of *malM::lacZ* fusion strains
(A) β-galactosidase activity measured in wild-type *S. pneumoniae, malM::lacZ fusion* strain (CH79) and *malM::lacZ*^{rpoB} fusion strain (CH80) grown to mid-log phase in either THY, THY sucrose or THY 0.8% maltose.
(B) β-galactosidase activity measured in CH80 was grown to mid exponential phase in different concentrations of THY 0.8% maltose.

levels over the wild-type background and still greater levels when grown in THY 0.8% maltose. The construct containing the *rpoB* RBS, CH80, gave very much greater activity than CH79 in all conditions tested, consistent with there being a greater baseline amount of enzyme when translational efficiency is improved. The β-galactosidase activity varied according to the maltose concentration in the medium (Figure 3.9). Given that the system appeared to allow some control over *lacZ* expression, a pneumococcal strain with the $tnpR^M$ allele inserted between *malM* and *malP*, CH102, on the chromosome and the antibiotic cassette present elsewhere on the chromosome was also constructed for use in assaying resolution.

3.2.8 Determination of resolvase production and resolution with $tnpR^M$

Western blots with anti-resolvase antibody were performed on total protein extracted from cells grown to mid log phase in THY or THY 0.8% maltose (Figure 3.10). Resolvase was detectable in the pneumococcal strain harbouring the *ply*:$tnpR^M$ fusion, CH104, and in the strain with $tnpR^M$ inserted between *malM* and *malP*, CH102. Therefore alteration of the codon sequence of the *tnpR* gene had resulted in improved mRNA translation and detectable levels of enzyme. To see if these improvements resulted in functional resolvase enzyme, enabling resolution in *S. pneumoniae*, resolution was assayed for in cells grown to mid log phase *in vitro* in both the *ply* fusion strain and the maltose inducible strain and *in vivo* in the *ply* fusion only. CH49, the *S. pneumoniae* strain harbouring the antibiotic cassette alone, was used as a control to assess the background level of spontaneous resolution in all experiments (Figure 3.11). There was no detectable spontaneous resolution in CH49, but resolution was detectable *in vitro* in CH104, the *ply* reporter strain, and CH102, the maltose inducible strain. The level of resolution in each strain was ~30%. There was no significant difference in the

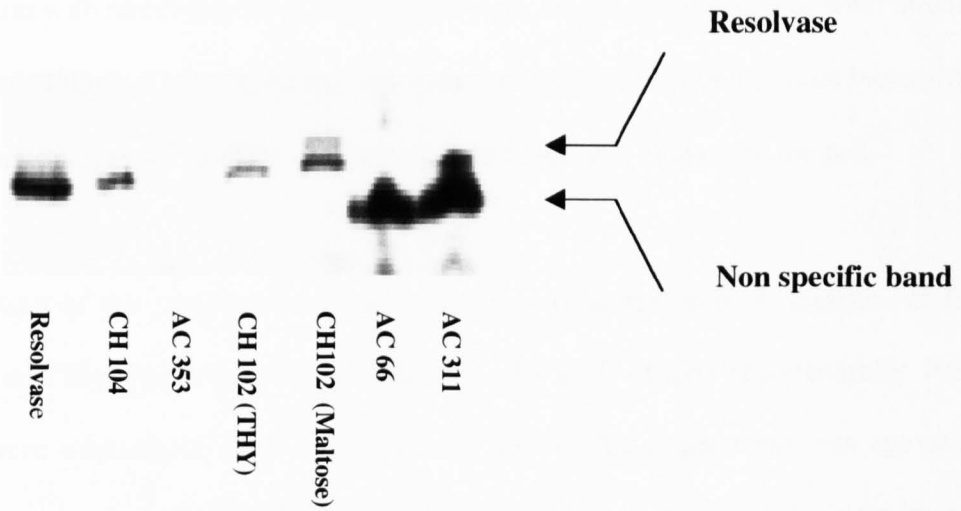


Figure 3.10 Western blot for Resolvase protein in *ply::tnpR^M* reporter strain (CH104) and *malM::tnpR^M* reporter strain (CH102).
Western blot probing total protein extracted from cells grown to mid exponential phase with anti-resolvase antibody; Lane 1, purified resolvase enzyme, lane 2, (CH104), *S. pneumoniae* strain harbouring *orf2::res1-aad9-res1*and *ply::tnpR^M*fusion, lane 3, (AC353) wild-type *S. pneumoniae*, lanes 4 and 5, (CH104), *S. pneumoniae* strain harbouring *orf2::res1-aad9-res1*and *malM::tnpR^M*fusion grown in either THY or THY 0.8% maltose, lane 6, (AC66)*V. cholerae* strain harbouring the *res-tet-res* cassette alone and lane7, (AC311) *V. cholerae* strain harbouring *res-tet-res* and *vieS'::tnpR* fusion. The same quantity of protein was loaded in each lane.

level of resolution with increasing maltose concentrations in either maltose-inducible strain studied although there appeared to be a trend towards decreasing amounts of resolution with increasing maltose concentration. This is the opposite of what might be predicted (Figure 3.11) suggesting that there was a disconnection between increasing transcription of *tnpR*^{M} and the concentration of functional enzyme in the cell.

Resolution of the *ply:tnpR*^{M} reporter strain was assessed in a murine model of pneumonia. Eight mice were inoculated with ~ 10^{7} CFU of CH102 intranasally. Pairs of mice were euthanased at 2, 8, 24 and 44 hours. The experiment was repeated in duplicate on a separate day. The percentage of Spc^{S} pneumococci in the input inoculum and those recovered at each time point was determined (>400 colonies were tested for each mouse). The amount of resolution was similar between the two experiments but was at a level 3 fold less than that seen *in vitro*. Maximum resolution was seen at 24 hours and was only in the order of 15%.

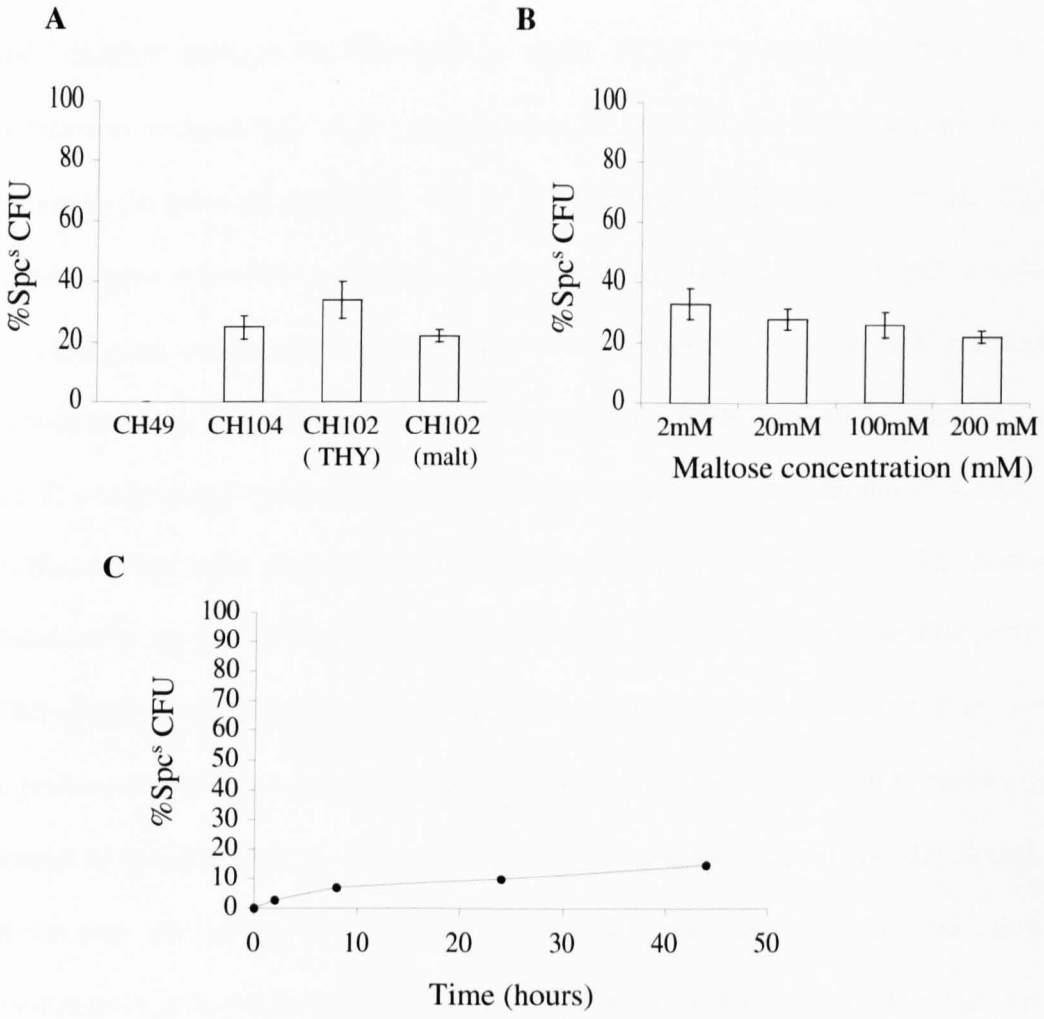


Figure 3.11 Assessment of resolution of reporter strains *in vitro* and *in vivo*
(A) *S. pneumoniae* strains with either the *res1-aad9-res1* cassette alone (CH49), *ply::tnpR*^{rpoB} fusion and cassette (CH104) or *malM:: tnpR*^{rpoB} fusion (CH102) were grown to mid log phase in THY broth or THY 0.8% maltose. The percentage of Spc^s CFU, shown on the Y axis, was determined for each strain.
(B) Strain CH102 was grown to mid log phase in varying concentrations of THY maltose. The percentage of Spc^s CFU was determined for each media.
(C) The kinetics of resolution as a marker for transcriptional induction of *ply* during *S. pneumoniae* infection. Eight Swiss-Webster mice were inoculated with1x10^7 CFU of CH104. Pairs of mice were euthanased at 2, 8, 24 and 48 hours post infection each and the percentage of Spc^s CFU in the lung at each time point was determined. The experiment was repeated in duplicate. Results of one experiment are shown.

3.3 SUMMARY

This chapter outlines the attempts to apply RIVET, recombinase-based *in vivo* expression technology, to *S. pneumoniae* in order to use the technique to study pneumolysin gene expression *in vivo* in a murine model of infection. Initial attempts failed to give detectable resolution *in vitro* or *in vivo*. It was demonstrated that though the *tnpR* gene was transcribed under *in vitro* conditions there was no protein detectable by western blot. Replacement of the native resolvase RBS with one identical to that of the *S. pneumoniae rpoB* gene with a view to improving translational efficiency was ineffective and there was still no detectable resolvase enzyme produced. The nucleotide sequence of the *tnpR* gene and upstream 5' region was altered to include this preferable RBS upstream of the ATG start codon and replace any codons that were rarely used in *S. pneumoniae* with a more commonly used ones. In parallel a system to enable tighter control of specific gene expression with use of the maltose operon was developed. Use of the new altered $tnpR^{M}$ allele led to detectable resolvase protein production and resolution *in vivo* and *in vitro*. However the level of resolution was not robust enough nor responsive to increasing levels of transcription to be able to use the system as it was to answer the initial questions that it had been developed for.

CHAPTER 4 CHARACTERISATION OF VIRULENCE GENE REGULATORS IN *S. PNEUMONIAE*

4.1 INTRODUCTION

The work described in the previous chapter was performed with the aim of developing a method to study virulence gene regulation in *S. pneumoniae in vivo*. To this end attempts were made to implement RIVET in *S. pneumoniae*. When it became apparent that this methodology was not going to be robust or sensitive enough in *S. pneumoniae* to use in further experiments I proceeded to investigate other methods of studying virulence gene regulation in this organism.

Three large scale STM screens in *S. pneumoniae* have been published (Hava and Camilli, 2002; Lau *et al.*, 2001; Polissi *et al.*, 1998). One study took 1786 insertion-duplication mutants to screen simultaneously for loss of virulence in both a murine model of pneumonia and one of bacteraemia (Lau *et al.*, 2001). Of these, 186 mutants were initially identified as attenuated. Sixty were attenuated in both models, 62 were attenuated solely in bacteraemia and the remaining 64 in pneumonia alone. Amongst the genes identified as being required for causing disease were some predicted to encode for transcriptional regulators or part of a signal transduction system. In a second study, 6149 *mariner* transposon insertion strains were screened. Of these, 387 mutants were attenuated for lung infection and 29 of these are thought to be involved in signal transduction or gene regulation, based on sequence similarity searches (Hava and Camilli, 2002). The phenotypes of a subset of the total 387 mutants were examined in two other murine models; bacteraemia and nasopharyngeal carriage. This identified four classes of mutants based on attenuation in animal infection models. Those

attenuated in causing (1) pneumonia, (2) pneumonia and bacteraemia, (3) pneumonia and nasopharyngeal carriage or (4) all three (Hava and Camilli, 2002). One interpretation of these data is that different virulence factors play specific roles in different tissues and/or at different times of the infective process. This would inevitably need the involvement of regulators that can orchestrate the overall changes in patterns of gene transcription, translation, protein activation, secretion etc. I hypothesized that genes identified in the STM screen that were predicted to encode proteins with a regulatory function might be involved in this virulence gene regulation. I planned to determine if this was the case by examining phenotypes of the mutants in different murine models of infection. I then intended to identify the genes in any 'regulon' with the use of transcriptional profiling.

4.2 RESULTS

4.2.1 Differential attenuation of regulator mutants in different infection models

On the basis of sequence similarity searches using the BLAST program (http://www.ncbi.nlm.nih.gov/blast/) and knowledge of neighbouring genes on the *S. pneumoniae* chromosome five of the genes identified by an STM screen (Hava and Camilli, 2002) were selected for study (*sp0156, sp0247, sp0461, sp0661 and sp1800).* The *magellan2* insertion mutations were backcrossed into AC353. In the cases of *sp0661* and *sp1800* further mutants were constructed as described below.

The genes *sp0156* and *sp0661* are predicted to encode response regulator components of a TCSTS and are adjacent to genes predicted to encode corresponding histidine kinase components (*sp0155* and *sp0662* respectively). Although both genes were identified in the STM screen it was possible that the attenuated phenotype of these transposon

mutants was due to a polar effect on other genes in the operon. The open reading frame of *sp0661* overlaps the open reading frame of *sp0662* and *sp0156* is only 51bp away from the predicted start site of the next open reading frame/gene *sp0157*. For this reason I wished to construct two new mutants with in-frame deletions in *sp0156* or *sp0661*. A *S. pneumoniae* strain, CH119, with an unmarked in-frame deletion of 529 bp following the ATG codon of *sp0661* was constructed but a strain with an unmarked in-frame deletion in *sp0156* could not be made despite multiple attempts and hence the *magellan2* insertion mutation in *sp0156* was backcrossed into AC353 and this strain was used in further competition assays.

To determine if strains with mutations in the 5 chosen genes with predicted regulator function had global virulence defects or if they showed more selective attenuation in certain types of disease, competition assays in murine models of pneumonia, bacteraemia and nasopharyngeal carriage were performed. Mice were inoculated intraperitoneally with ~ 1 x 10^6 bacteria for the bacteraemia model, with ~ 1 x 10^7 intranasally for the pneumonia model and with ~ 1 x 10^8 intranasally for the nasopharyngeal carriage model. Bacteria were recovered from the blood, lungs or nasopharyngeal washes respectively at 22 hours, 44 hours or 7 days post-inoculation. The *in vivo* competition index (CI) was calculated for each mouse by dividing the ratio of the mutant to wild-type bacteria, that were recovered from the mouse, by the ratio of the mutant to wild-type bacteria that were inoculated into each animal. The geometric mean of the CIs for each strain was determined and is listed in Table 4.1. A mean CI of less than 1 indicates a defect in virulence of the test strain as compared to the wild-type.

TABLE 4.1. Virulence in mice of regulator mutant strains in three murine models of infection assessed by competition assays

Strain	Relevant Genotype	Geometric mean *in vivo* CI[a]		
		Lung infection	Carriage	Bacteraemia
CH108	AC353 *sp0156::magellan2* Sm^R Cm^R	**<0.008 (6)**	0.4 (8)	1.7 (4)
CH107	AC353 *sp0247::magellan2* Sm^R Cm^R	**<0.08 (4)**	**<0.09 (7)**	**0.25 (4)**
STM90[b]	AC353 *sp0661::magellan2* Sm^R Cm^R	**<0.017 (4)**	ND	0.55 (7)
CH119	AC353 Δ*sp0661* Sm^R	**0.004 (6)**	0.75 (6)	**<0.001 (4)**
AC1213	AC353 *rlrA::magellan2* Sm^R Cm^R	**<0.30 (7)**	**<0.071 (10)**	0.74 (4)
STM206	AC353 *mgrA 5' UTR::magellan2* Sm^R Cm^R	**<0.023 (4)**	**0.009 (6)**	0.56 (3)
AC1272	AC353 *mgrA::magellan2* Sm^R Cm^R	**0.47 (8)**	**0.58 (6)**	ND
AC1500	AC353 Δ*mgrA* Sm^R Spc^R	**0.26 (12)**	**0.034 (6)**	**2.1 (8)**

Table 4.1 [a] The geometric mean of the *in vivo* competition index (CI) is shown. The number of animals infected in each experiment is indicated in parentheses. In the competitions in which no mutant bacteria were recovered from a particular animal the number 1 was substituted as the numerator when determining the *in vivo* ratio for that animal, thus the *in vivo* mean CI is denoted as less than the calculated value. Geometric means shown in bold were considered significant (*P*-values < 0.05 by the student's two-tailed *t*-test). Open reading frames *sp0461*and *sp1800* are renamed *rlrA* and *mgrA*. [b] The competition assays with STM90 were performed by D. Hava. (ND not done)

4.2.2 RlrA and MgrA have similarity to transcriptional regulators of group A streptococcus (GAS)

What are these regulators regulating? I decided to focus more specifically on two regulators. *Sp0247* was excluded on the basis of the mutant having a global virulence defect suggestive that this transcriptional regulator could be involved in regulating factors required for basic cellular processes important in cell survival *in vivo* and these may not be differentially regulated *in vivo*. Given the difficulty in constructing a deletion mutant, *Sp0156* was also excluded. On the basis of sequence similarity searches using the BLAST program (http://www.ncbi.nlm.nih.gov/blast/) *sp0461* and *sp1800* were selected for further study.

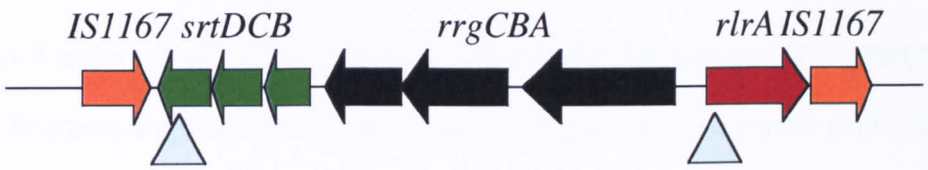
The mutants with transposon insertions in genes *sp1800* and *sp0461* were attenuated in models of pneumonia and nasopharyngeal carriage suggesting that the products of these two genes are required for interactions with the mucosal surfaces of the respiratory tract. Both of these genes are predicted to encode proteins with similarity to transcriptional regulators in GAS (*S. pyogenes*). *Sp0461* could encode a protein with 49% similarity to RofA and Nra (Fogg *et al.*, 1994; Podbielski *et al.*, 1999) and *sp1800* could encode a protein that is 51% similar and 25% identical to the multiple gene regulator, Mga (McIver and Scott, 1997; McIver and Myles, 2002). These two gene products in *S. pneumoniae* were renamed RlrA (RofA-like regulator A) and MgrA (Mga-like repressor A).

Mga and RofA/Nra of *S. pyogenes* regulate transcription of virulence genes encoding primarily surface-associated products. RofA and Nra have been shown to activate or

repress the transcription of genes encoding important adhesins such as fibronectin-binding protein (*prtF*), collagen binding adhesin (*cpa*) and a second fibronectin-binding protein (*prtF2*). Mga activates the transcription of numerous virulence genes involved in *S. pyogenes* pathogenesis, including the genes encoding M family of proteins (*emm, mrp, arp* and *enn*), C5a peptidase (*scpA*), serum opacity factor (*sof*) and secreted inhibitor of complement (*sic*) (Caparon and Scott, 1987; Chen *et al.*, 1993; Kihlberg *et al.*, 1995; McLandsborough and Cleary, 1995).

The *rofA* homologue *rlrA* lies in a 12 kb region of DNA that might constitute a pathogenicity islet and is divergently transcribed from six genes (*sp0462 – sp0468*). The islet is flanked by two IS1167 insertion sequences. The gene organization is shown in Figure 4.1. Three of the islet genes encode proteins with similarity to microbial surface components recognizing adhesive matrix molecules (MSCRAMMS) and three proteins similar to sortases (Figure 4.1). Sortases are enzymes that catalyse the covalent linkage of secreted proteins containing an LPXTG motif to the bacterial cell wall (Mazmanian *et al.*, 1999; Mazmanian *et al.*, 2001). Each of the putative MSCRAMMS, renamed RrgA, RrgB and RrgC have C-terminal sorting signals that are characteristic of LPXTG-containing proteins, except that the LPXTG sequence in each varies by two amino acids (Figure 4.1). One of the sortase genes, *srtD*, was also identified in the same STM screen that had identified *rlrA* as being required for lung infection and subsequent work showed that an *rrgA* mutant strain was attenuated in a pneumonia and nasopharyngeal carriage model (Hava and Camilli, 2002).

A



B

Protein	C-terminal sequence
RrgA	**YPRTG**GIGMLPFYLIGCMMMGGVLLYTRKHP
RrgB	**IPQTG**GIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQLA
RrgC	**VPDTG**EETLYILMLVAILLFGSGYYLTKKPNN

Figure 4.1 The *rlrA* locus

(A) The *rlrA* gene is divergently transcribed from six genes. The entire locus is flanked by one frame-shifted and one intact IS1167 element. The sites of *magellan2* insertions in *rlrA* and *srtD,* identified by STM, are shown with blue triangles.

(B) The predicted C-terminal sorting signals of the three MSCRAMMS, RrgA, RrgB and RrgC, are listed. Each is characteristic of proteins anchored to the cell wall by sortases. The LPXTG motif (in bold) varies by the first amino acid. This is followed by a stretch of hydrophobic residues (underlined) and a charged tail.

On the GAS chromosome *mga* lies directly upstream and reads in the same direction as some of the genes that it regulates including those encoding M protein and C5a peptidase (McIver and Myles, 2002; Pobielski *et al.*, 1996; Rasmussen *et al.*, 2000; Terao *et al.*, 2001). In contrast, the arrangement around *mgrA* on the *S. pneumoniae* serotype 4 genome is different. It is approximately 450 bp upstream and divergently transcribed from a cluster of five small open-reading frames that appear to encode hypothetical proteins of unknown function and downstream of a gene encoding a putative LacI family transcriptional regulator (*sp1799*) (Figure 4.2). The *mgrA* gene itself has a predicted transcriptional terminator directly after the stop codon. The region encoding MgrA was actually identified twice in the serotype 4 *S. pneumoniae* STM screen (Hava and Camilli, 2002). One transposon insertion mapped to the middle of the 1479 bp open-reading frame (AC1272) and a second to a non-coding region 300 bp upstream of the predicted start site of *mgrA* (*mgrA* 5' UTR or STM206) (Figure 4.2).

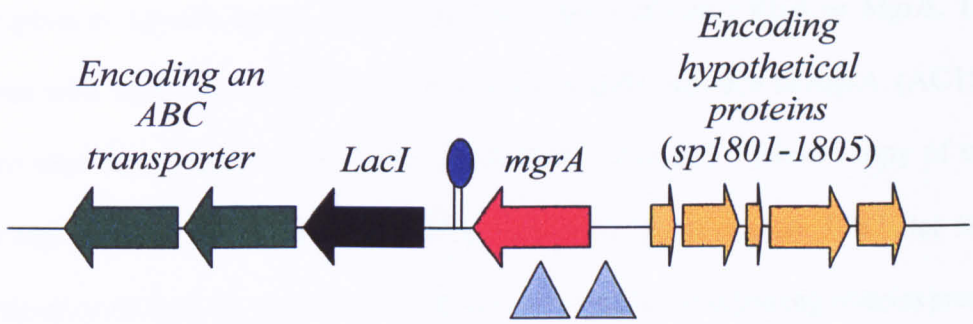


Figure 4.2 The *mgrA* locus

The *mgrA* (*sp1800*) gene is upstream and divergently transcribed from a cluster of five small open-reading frames (yellow) encoding putative hypothetical proteins of unknown function and upstream of a gene encoding a putative LacI family transcriptional regulator (*sp1799*). The *mgrA* gene itself has a predicted transcriptional terminator directly after the stop codon. The sites of *magellan2* insertions in or around *mgrA* identified by STM are shown with blue triangles.

These two STM mutants, AC1272 and STM206, were attenuated to different degrees in competition assays (Table 4.1). Though both were attenuated for pneumonia and carriage, AC1272 showed a lower level of attenuation. This strain has a *magellan2* transposon inserted at approximately base pair 750 in the open reading frame. The DNA binding activity of the Mga of GAS has been mapped to two helix-turn-helix regions in the first 140 amino acids from the N-terminus. It is possible therefore that a truncated but partially functional protein is made in the *S. pneumoniae* strain, AC1272, which has some effect on the virulence phenotype. To avoid this problem a deletion insertion mutant was constructed as described below.

4.2.3 Construction of an *mgrA* deletion strain and a strain over-expressing *mgrA* or *rlrA*.

Three further mutants were constructed to use in experiments assessing whether transcription of specific genes were under the control of either RlrA or MgrA. These were one with a deletion-insertion of a Spc^r gene, (*aad9*), in place of *mgrA*, (AC1500), and two merodiploid strains (AC1278 and AC1481) that had a second copy of either *rlrA* or *mgrA* inserted between *malM* and *malP* of the maltose operon. The latter strains place *mgrA* or *rlrA* under the control of the maltose promoter allowing over-expression of either gene product when grown in maltose (Figures 4.3 and 4.4). The phenotype of the new *mgrA* deletion mutant was also assessed in competition assays against the wild-type strain in the three murine models of infection. Similar to the transposon insertion mutants, the deletion strain was attenuated in both nasopharyngeal carriage and pneumonia but not in the bacteremia model (Table 4.1).

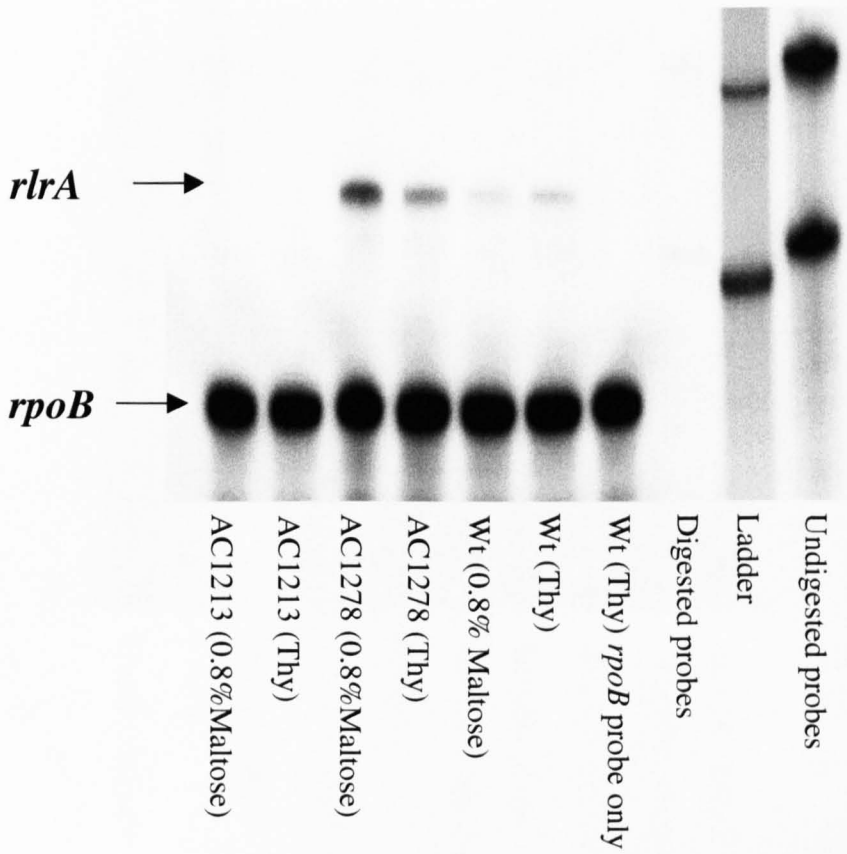


Figure 4.3 RPA for *rlrA* gene transcript

RPA was used to confirm over-expression of *rlrA* gene under the control of the maltose inducible promoter. RNA was harvested from cells grown to OD_{600} 0.4 in either THY or THY 0.8% maltose and probed for *rlrA*. A *rpoB* probe was used as a loading control. The *rlrA* transcript level is 7-fold greater in AC1278 grown in maltose than in wild-type cells.

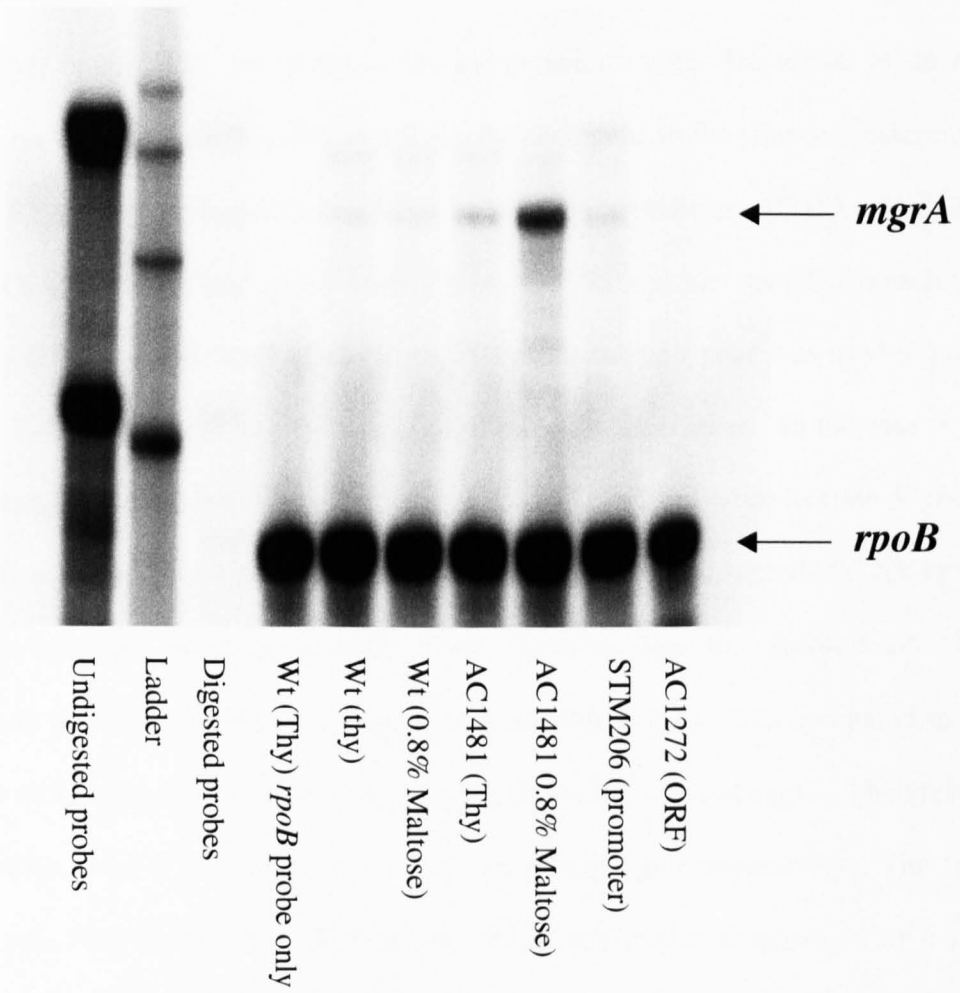


Figure 4.4 RPA for *mgrA* gene transcript

RPA was used to confirm over expression of *mgrA* gene under the control of the maltose inducible promoter. RNA was harvested from cells grown to OD_{600} 0.4 in either THY or THY 0.8% maltose and probed for *rlrA*. A *rpoB* probe was used as a loading control. The *mgrA* transcript level is 12-fold greater in AC1481 grown in maltose than in wild-type cells. Two *mgrA*-specific mRNA species were detected with the riboprobe.

4.2.4 Each gene in the pathogenicity islet requires RlrA for expression

On the basis of the organization of the islet and similarity to RofA it was hypothesized that *rlrA* acts as a transcriptional regulator that activates its own transcription and as well as that of the other six genes in the pathogenicity islet. The effect of an *rlrA* mutation on the steady-state levels of mRNA for each gene in the islet was determined by RPA. RPAs were performed using RNA isolated from wild-type AC353 or AC1213, the strain that harbors a transposon insertion in *rlrA*. Riboprobes specific to each islet gene, as well as to *rpoB*, were synthesized. The probe to *rpoB* gene was used to probe the same RNA and served as a loading control for each experiment. In the case of the seven genes within the islet, the riboprobes were designed to the immediate 5' end of each ORF, overlapping the putative translation initiation site, and such that a 200 to 400 nucleotide fragment would be protected from digestion from the ribonuclease. The steady-state level of mRNA of each gene was decreased in AC1213 compared to the wild-type strain (Figure 4.5). The level of decrease varied with each gene. The greatest decrease (10- and 11-fold) was observed for *rrgB* and *rrgC*, respectively. The *rrgA* message was only decreased by 2.5-fold and *srtB*, *srtC* and *srtD* messages by 6-, 7-, and 8-fold, respectively. The variability in detectable transcript level between these six genes is consistent with the presence of three promoters controlling their transcription: one for *rrgA* a second *rrgB* or *rrgC* and a third controlling *srtB*, *srtC* and *srtD*. The *srtB* probe protected three differently sized messages, suggesting the possibility that there are multiple transcriptional start sites within the sequence of the riboprobe. Each promoter appeared to be activated by RlrA. DNA binding studies and primer extension analysis, performed by Dave Hava, mapped the transcriptional start sites of promoters upstream of *rlrA*, *rrgA*, *rrgB*, and *srtB* and

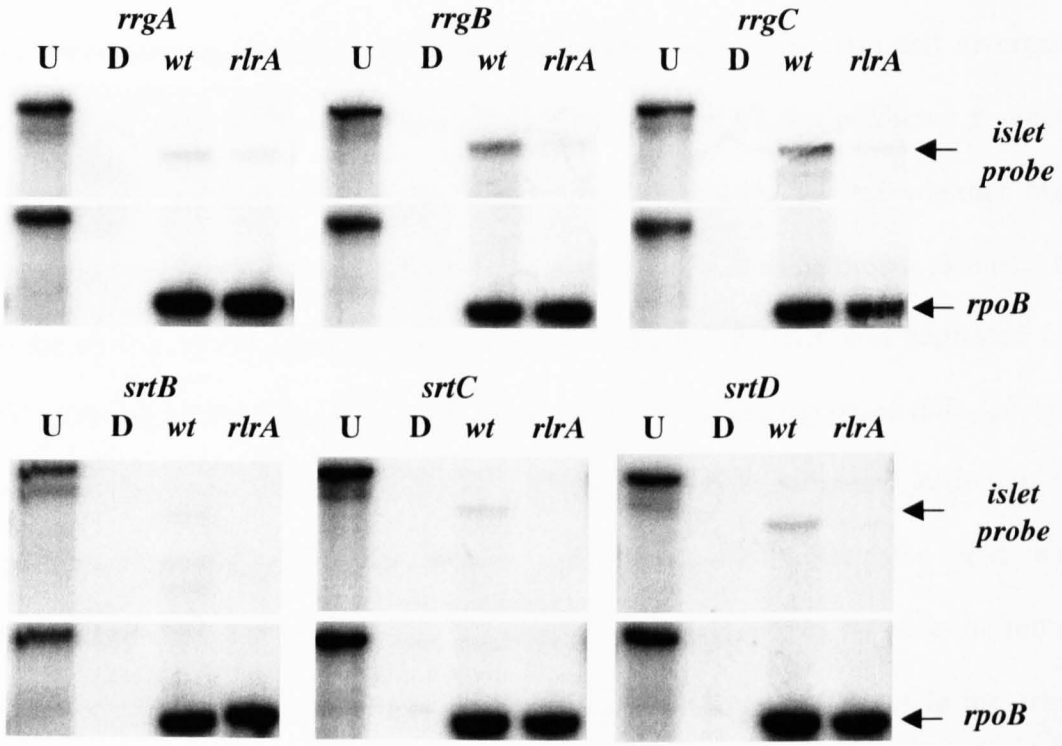


Figure 4.5 RPAs of *rlrA* islet gene transcripts

RPA to examine the expression of each of the *rlrA* islet genes in wild-type (wt) compared to *rlrA*::*magellan2* transposon insertion strain *(rlrA)*. RNA was harvested from cells grown to OD_{600} 0.4 in THY. A *rpoB* probe was used as a loading control. The results shown are from an experiment performed by Dave Hava.

confirmed that RlrA bound to the predicted promoter sequences (Hava *et al.*, 2003a).

4.2.5 Unlike *mga*, *mgrA* does not appear to regulate neighbouring genes.

As previously mentioned *mgrA* is approximately 450 bp upstream and divergently transcribed from a cluster of five small open-reading frames predicted to encode hypothetical proteins of unknown function (Figure 4.2). To test whether MgrA regulates transcription of this cluster RPAs were performed using probes against 2 orfs in the cluster, *sp1803* and *sp1804*. In these experiments RNA was harvested from AC1500 and AC1481 grown in 0.8% maltose to mid exponential phase (OD_{600} 0.4). A probe to *rpoB* gene served as a loading control for each experiment. Although both *sp1803* and *sp1804* were transcriptionally active, there was no detectable difference in the level of either transcript in the *mgrA* over-expressing versus the deletion mutant. Therefore *mgrA* does not appear to affect transcription of these genes in the growth conditions tested (Figure 4.6).

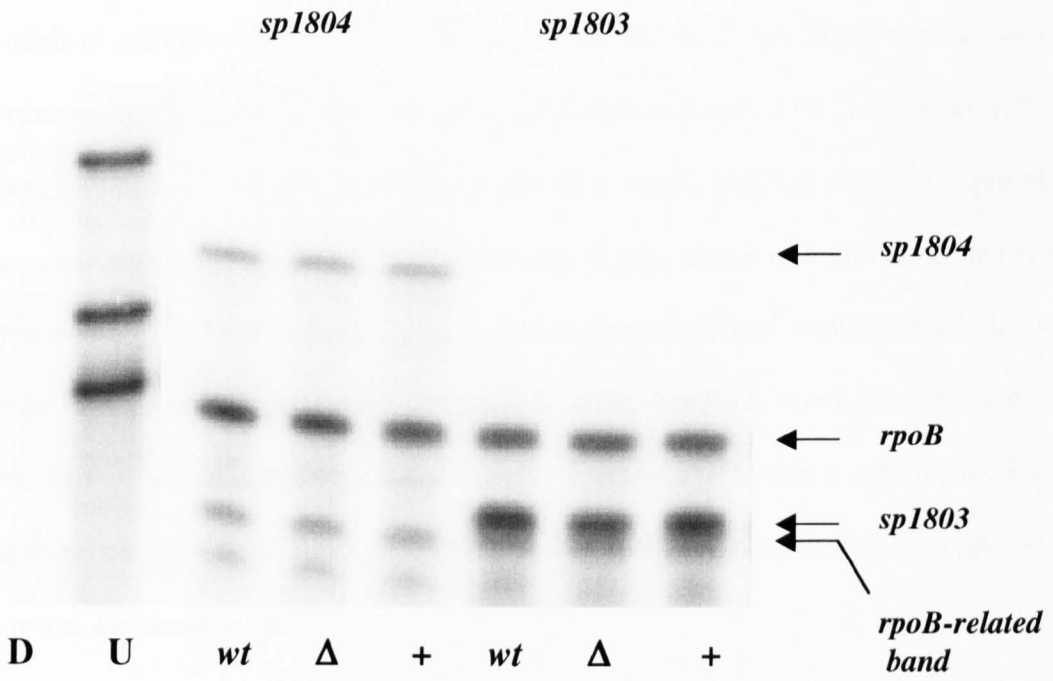


Figure 4.6 RPA to analyze the mRNA levels of *sp1803* and *sp1804* in wild-type (*wt*; AC353), *mgrA* deletion strain (Δ; AC1500) and *mgrA* over-expressing strain (+; AC1481). Riboprobes to *sp1803, sp1804* and *rpoB* were generated and hybridized to 10μg of *S. pneumoniae* RNA from the three strains. RNA was harvested from cells grown in 0.8% maltose to OD 0.4. Lanes U and D contain undigested riboprobes and riboprobes digested by Rnase in the absence of *S. pneumoniae* RNA.

4.3 SUMMARY

This chapter described the analysis of the virulence defect of 5 putative regulator mutants that were identified in an STM screen of TIGR4 assessed by using three murine models of infection. Two of these putative regulators, RlrA and MgrA were chosen for further examination. Both were similar to transcriptional activators of virulence genes in GAS. Given this similarity and the fact that they were identified as being required for nasopharyngeal carriage and for development of pneumonia in a murine model it was hypothesised that they would be involved in transcriptional regulation of virulence genes in *S. pneumoniae*. It was confirmed by RPA that RlrA acted as a transcriptional activator of six neighbouring genes. One of these, *srtD*, was a previously known virulence factor. MgrA, however, did not appear to effect transcription of neighbouring genes as assessed by RPA.

CHAPTER 5 INTERPLAY BETWEEN MGRA AND THE *rlrA* PATHOGENICITY ISLET

5.1 INTRODUCTION

MgrA has significant similarity to Mga, an important virulence gene regulator in GAS. Deleting *mgrA* in *S. pneumoniae* resulted in an attenuated phenotype showing that it is likely to be an important regulator of virulence genes in this pathogen. Preliminary experiments, described in the previous chapter, to identify some of the possible targets of this putative transcriptional regulator suggested that, at least *in vitro*, MgrA is not required for and does not regulate transcription of a neighbouring cluster of genes on the chromosome.

The TIGR4 chromosome contains 2236 predicted coding regions of which approximately 65% have been assigned a biological role (Tettelin *et al.*, 2001). Theoretically any of these 2236 genes could be a target for MgrA. At the time of conception of this project high-density DNA microarrays had been applied to a wide variety of systems for comparative genomic analysis (Behr *et al.*, 1999), to monitor gene expression in whole genomes (deLisa *et al.*, 2001) and more specifically for the identification of regulons within *S. pneumoniae* (de Saizieu *et al.*, 2000). This chapter describes work performed to identify further targets of MgrA with the aid of microarray analysis. Since starting this work there have been a number of reports describing microarray technology to identify genome wide changes in expression of genes in *S. pneumoniae* strains with mutations in putative TCSTS or transcriptional regulators (Blue and Mitchell, 2003; Sebert *et al.*, 2002).

5.2 RESULTS

5.2.1 MgrA acts as a transcriptional repressor of the *rlrA* pathogenicity islet.

The GAS homologue, Mga, shows variable transcription and activity at different stages in the growth phase (McIver and Scott, 1997). It is transcribed during exponential phase but transcription ceases as the bacteria enter stationary phase. To assess if the transcription of *mgrA* varied with growth RPAs were performed to analyse mRNA levels of the *mgrA* gene in AC353 at the four different time points during growth approximating to early, mid, late exponential and early stationary phase (Figure 5.1). Assessment of mRNA levels at late stages of stationary phase in *S. pneumoniae* is hindered by the fact that bacterial cells undergo autolysis and bulk mRNA degradation at this time (data not shown). Two mRNA species were detected with the *mgrA* riboprobe at all 4 time points. This is consistent with the gene having two transcriptional start sites as the riboprobe for *mgrA* was complementary to RNA overlapping the 5' untranslated mRNA and the coding sequence of the gene. MgrA transcript levels did not change over exponential phase to early stationary phase of growth. However post-transcriptional regulation and/or changes in the activity of the translated protein cannot be excluded.

cDNA microarray analysis of gene expression patterns of the *mgrA* deletion and over-expressing strains (AC1500 and AC1481, respectively) was used in an attempt to identify genes regulated by MgrA. On the hypothesis that, similar to Mga, MgrA may not be active at all phases of growth, RNA for use on the microarrays was harvested from all 4 time points in the growth curve (as above). All three strains had similar growth kinetics *in vitro* (Figure 5.1). The vast majority of genes displayed comparable expression levels between the two strains (Figure 5.2).

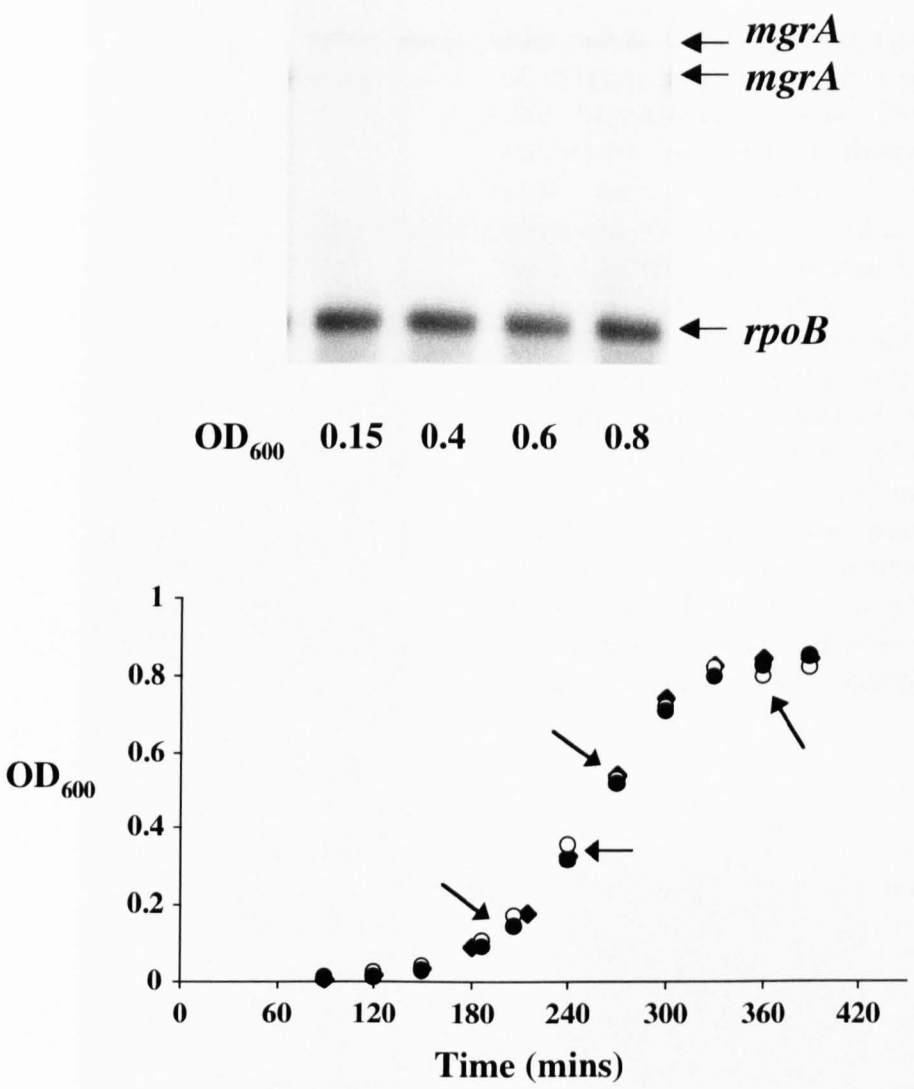


Figure 5.1 RPAs of *mgrA* transcript and comparison of growth *in vitro*
(A) RPAs were performed to analyse mRNA levels of the *mgrA* genes in AC353 at the four different time points. Riboprobes to *mgrA* and *rpoB* were generate and hybridised to 5 µg of total *S. pnuemoniae* RNA (AC353). Samples are loaded as follows; lane 1 undigested probe alone and lanes 2 to 5 probes hybridised to RNA harvested from cells grown to the OD_{600} indicated.
(B) Growth curves for the wild-type (AC353 ♦), *mgrA* deletion mutant (●) and *mgrA* over-expressing strain (○) cultured statically in Todd Hewitt broth plus 5% yeast extract (THY) and supplemented with oxyrase (5µl ml^{-1}) and 0.8% maltose. Aliquots were removed for RNA extraction at 4 time points, indicated by the arrows, (equivalent OD_{600} values: early log ~ 0.2, mid log ~ 0.4, late log ~ 0.6 and stationary phase ~ 0.8)

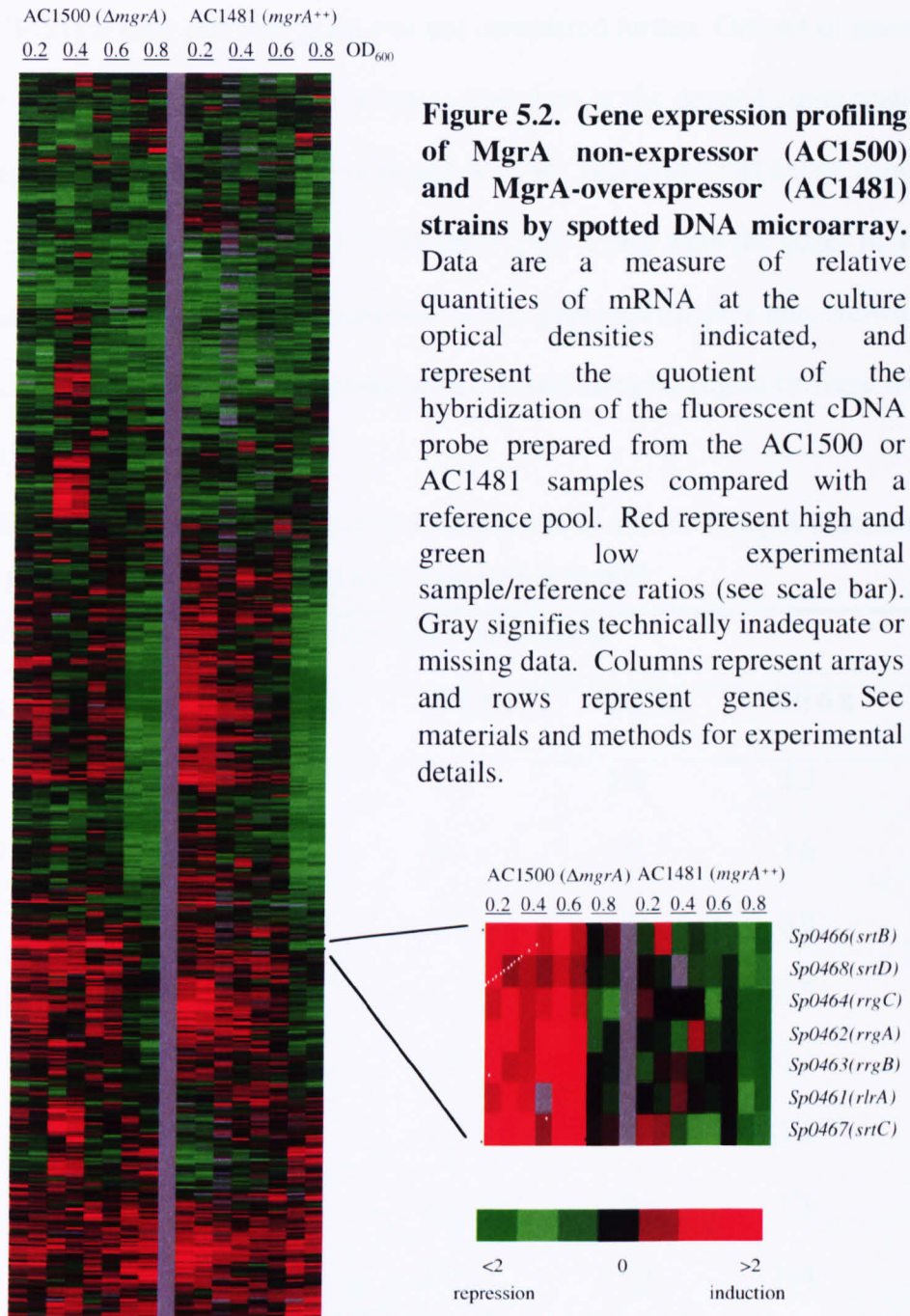


Figure 5.2. Gene expression profiling of MgrA non-expressor (AC1500) and MgrA-overexpressor (AC1481) strains by spotted DNA microarray. Data are a measure of relative quantities of mRNA at the culture optical densities indicated, and represent the quotient of the hybridization of the fluorescent cDNA probe prepared from the AC1500 or AC1481 samples compared with a reference pool. Red represent high and green low experimental sample/reference ratios (see scale bar). Gray signifies technically inadequate or missing data. Columns represent arrays and rows represent genes. See materials and methods for experimental details.

A subset of genes that showed altered profiles between the two mutant strains (AC1500 and AC1481) at only one time point was not considered further. One set of genes were notable in that their transcripts were more abundant in the Δ*mgrA* strain than *mgrA* over-expressing strain at all four time points assessed, suggesting that MgrA is involved in repressing these genes (Figure 5.2 and Table 5.1). These were the genes in the *rlrA* pathogenicity islet which was discussed in the previous chapter and shown to be virulence factors required for both lung infection and nasopharyngeal carriage (Hava *et al.*, 2003a).

Table 5.1. Average mRNA level of *rlrA* islet genes in AC1500 (*mgrA* deletion strain) relative to AC1481[a] (*mgrA* over-expressing strain)

Fold difference in the mRNA level across the time course

TIGR4 ORF[b]	Gene	OD 0.2	OD 0.4	OD 0.6	OD 0.8
SP0461	*rlrA*	2.0	1.6	2.8	2.3
SP0462	*rrgA*	2.4	2.1	3.0	3.6
SP0463	*rrgB*	2.2	1.9	3.5	2.8
SP0464	*rrgC*	1.8	1.7	3.1	2.5
SP0466	*srtB*	2.0	2.8	2.7	2.8
SP0467	*srtC*	1.7	2.7	4.4	2.4
SP0468	*srtD*	2.4	1.7	2.3	1.5
SP1800	*mgrA*	10.2	9.7	15.6	NA

[a] The $\log_2$ red/green ratio values from replicate arrays were averaged at each time point. For each OD sample, the averaged $\log_2$ data from each AC1481 sample was subtracted from the respective $\log_2$ AC1500 data. The resulting data is expressed as the fold difference in levels between the two strains. Data from all arrays used in this study are also available (Gollub *et al.*, 2003) (Hemsley *et al.*, 2003). NA - not available.

[b] Designations of the ORFs in the TIGR4 sequence which are on the microarray.

Confirmation of these results was obtained by RPAs performed with RNA isolated from three independent sets of cultures and assessing two time points in the growth curve for each (mid exponential phase, OD_{600} 0.4, and late exponential phase OD_{600} 0.6). The islet contains four promoters driving transcription of seven genes (Hava and Camilli, 2002; Hava *et al.*, 2003a). Probes to a gene in each transcriptional unit, namely *rlrA, rrgA, rrgB* and *srtD* as well to *rpoB* were synthesized. As in previous RPAs, *rpoB* was used as a loading control for each experiment. All four of the tested islet genes showed increased mRNA levels in the *mgrA* deletion strain and decreased mRNA levels in the *mgrA* over-expressing strain at both time points and to a similar degree to that revealed by the microarray experiments (Figure 5.3). The level of each islet gene transcript was increased 2 – 3 fold in the *mgrA* deletion strain and decreased at least 2 fold in the *mgrA* over-expressing strain relative to wild-type at all time points tested. Together these data are consistent with MgrA acting as a transcriptional repressor of the islet genes. MgrA might bind directly to each promoter or alternatively could function indirectly by altering the expression of *rlrA*, a transcriptional activator, which subsequently effects the expression of the other six islet genes.

5.2.2 RlrA does not regulate *mgrA* transcription

RlrA is similar to RofA and Nra of GAS (Hava *et al.*, 2003a). In GAS Nra acts as a transcriptional repressor of *mga* whilst in turn Mga activates *nra* transcription (Pobielski *et al.*, 1999). To test the possible role of RlrA in regulation of *mgrA* in *S. pneumoniae* serotype 4 transcript levels of *mgrA* in wild-type and strains over-expressing *rlrA* or not expressing *rlrA* were assessed. RPAs were performed, probing for *mgrA*, using RNA harvested from wild type AC353, AC1213 (a strain that habours a transposon insertion in *rlrA*) and AC1278 (a strain that over-expresses *rlrA* from the

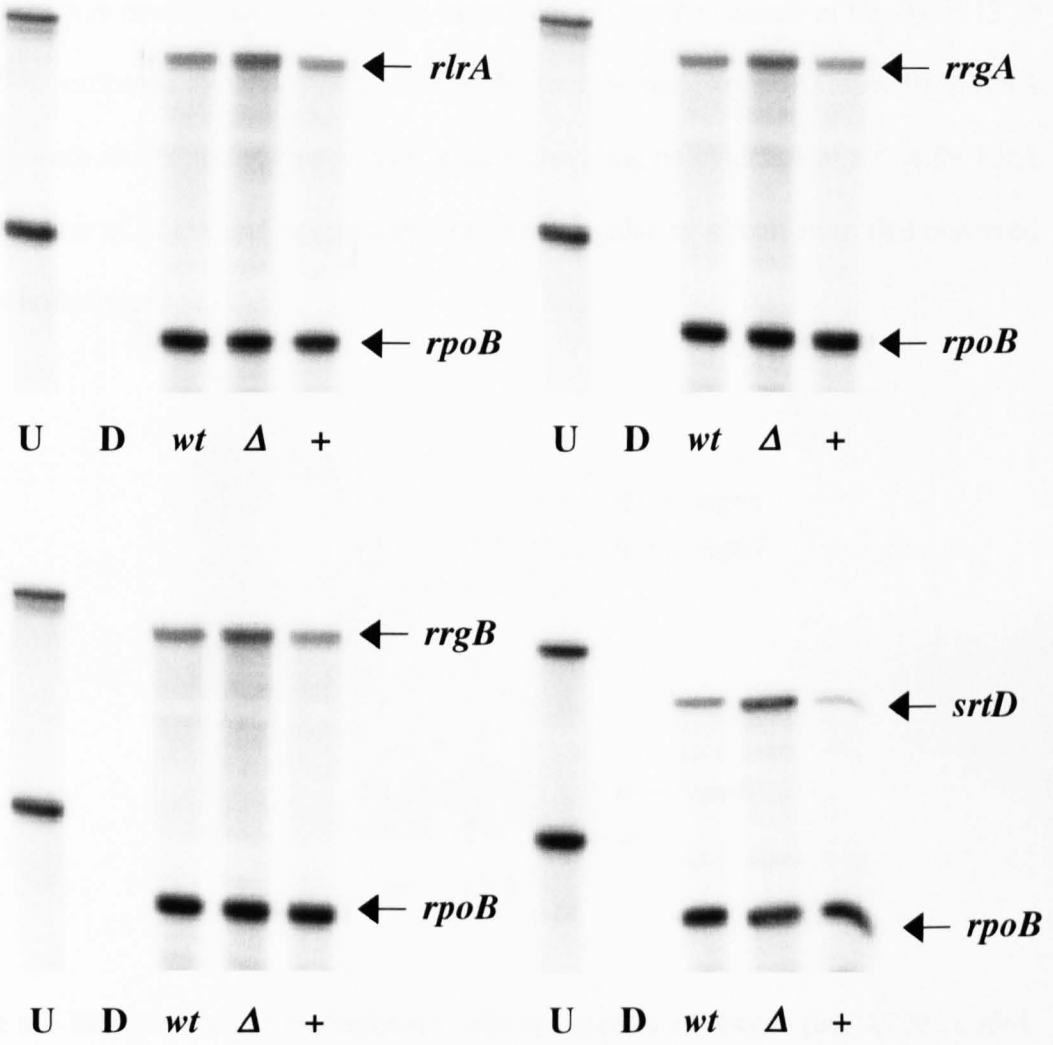


Figure 5.3 RPAs on *rlrA* islet genes

RPAs were performed to analyze the mRNA levels of a gene in each transcriptional unit in the *rlrA* islet, in the wild-type (*wt*; AC353), *mgrA* deletion strain (Δ, AC1500) and *mgrA* over-expressing strain (+, AC1481). Riboprobes to each gene, as well as *rpoB*, were generated and hybridized to 10μg of *S. pneumoniae* RNA from the three strains. RNA was harvested from cells grown in 0.8% maltose to OD 0.4 or 0.6. Transcript levels were comparable at either time point. Results from an OD 0.4 are shown. Lanes U and D contain undigested riboprobes and riboprobes digested by Rnase in the absence of *S. pneumoniae* RNA.

maltose inducible promoter). *RpoB* was used as a loading control. As shown in Figure 5.4, there was no detectable difference in amount of *mgrA* message in the AC1213 or AC1278 compared to wild-type (AC353). Neither of the two *mgrA* specific mRNA species were altered in the *rlrA* mutant strains providing no evidence of a role for RlrA in regulation of MgrA and suggesting a different regulatory circuit from that observed for the homologues in GAS.

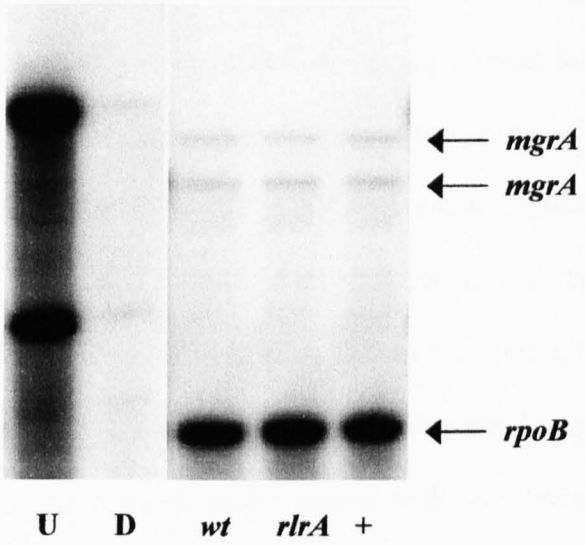


Figure 5.4 RPA to analyze the mRNA levels of *mgrA* in wild-type (*wt*; AC353), *rlrA* transposon insertion strain (*rlrA*; AC1213) and *rlrA* over-expressing strain (+; AC1278). Riboprobes to *mgrA* and *rpoB* were generated and hybridized to 10 μg of *S. pneumoniae* RNA from the three strains. RNA was harvested from cells grown in 0.8% maltose to OD 0.4. Lanes U and D contain undigested riboprobes and riboprobes digested by RNase in the absence of *S. pneumoniae* RNA.

5.2.3 ZmpB operon

Amongst the subset of genes that showed consistently altered profiles between the two mutant strains at only two time points in the arrays studies were *sp0663* and *sp0668*. These are predicted to be part of a putative operon, containing amongst other genes,

sp0661, encoding the response regulator component of TCS09 (nomenclature of Lange et al, (Lange *et al.*, 1999)), *sp0663* (encoding a conserved hypothetical protein) and *sp0664* (encoding a zinc metalloproteinase ZmpB), all of which had been identified by STM as being required for lung infection (Hava and Camilli, 2002). *Sp0663* and *sp0668* showed a 1.5 and 4 fold increase in mRNA levels in the *mgrA* over-expressing strain as compared to the deletion strain at the two later time points in the growth curve. The other genes in the putative *zmpB*-operon did not show similar changes. Given that more than one gene in the region was identified as having altered mRNA levels on the microarray screen and that the region was suspected to contain at least one virulence factor, RPAs were performed to determine if this apparent up-regulation by MgrA could be confirmed. RNA was isolated from three independent sets of cultures. The last two time points in the growth curve for each (late exponential phase OD_{600} 0.6 and early stationary phase OD_{600} 0.8) were examined. Probes to *sp0663* and *sp0668* as well to *rpoB* were synthesized. In contrast to the results with the microarray there was no detectable difference in amount of *sp0663* or *sp0668* message in the AC1500 or AC1481 compared to wild-type (Figure 5.5). Since the array results could not be confirmed by an alternative method this was not pursued further.

5.2.4 RlrA locus is not present in all strains that contain mgrA

Comparative genomics between TIGR4 and D39/R6, a serotype 2 *S. pneumoniae* strain, reveals that none of the *rlrA* islet genes are present in the latter strain but the *mgrA* gene is present, *spr1622* (Tettelin *et al.*, 2001). There is also a second gene (*spr1404*) in R6 with sequence similarity to *mga* and *mgrA*. It is 40.5% identical and 61.1% similar to the TIGR4 *mgrA* gene (Figure 5.6).

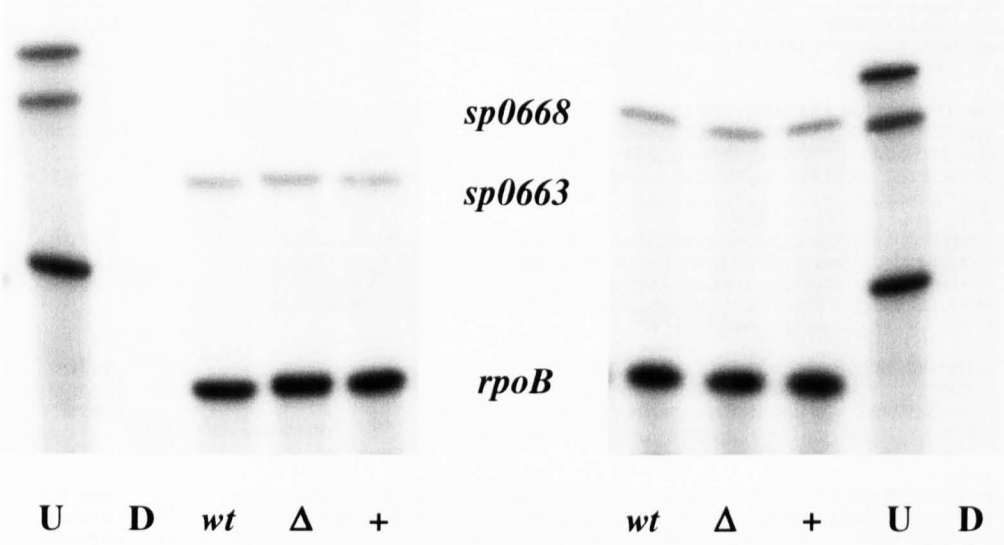


Figure 5.5 RPAs on *zmpB* operon gene transcripts *sp0663* and *sp0668*
RPAs were performed to analyze the mRNA levels of *sp0663* and *sp0668* in the wild-type (*wt*; AC353), *mgrA* deletion strain (Δ, AC1500) and *mgrA* over-expressing strain (+, AC1481). Riboprobes to each gene, as well as *rpoB*, were generated and hybridized to 10μg of *S. pneumoniae* RNA from the three strains. RNA was harvested from cells grown in 0.8% maltose to OD 0.6 or 0.8. Transcript levels were comparable at either time point. Results from an OD 0.6 are shown. Lanes U and D contain undigested riboprobes and riboprobes digested by Rnase in the absence of *S. pneumoniae* RNA.

Figure 5.6 Alignment of the amino acid sequences of Sp1800 (MgrA), Spr1622, Spr1404 and Mga

The amino acid sequences of proteins MgrA from *S. pneumoniae* TIGR4 and Spr1622 and Spr1404 from *S. pneumoniae* R6 were aligned with the sequence of Mga from *S. pyogenes*. An asterix (*) indicates identical residues in all four proteins, a semi colon (:) identical or similar residues in three and a single dot (.) identical or similar residues in two of the proteins.

To look further for the presence of *mgrA* and *rlrA* across serotypes, dot-blot hybridizations were performed on chromosomal DNA from eight clinical isolates of different serotypes using a control probe to *rpoB* and probes to *mgrA* and *rlrA*. The *rpoB* probe was used as a control. The same strains were also tested by PCR for the presence of all three genes and two other islet genes, *srtD* and *rrgA*. The serotypes screened were chosen on the basis that they represent all 7 serotypes covered by the pneumococcal conjugate vaccine, Prevnar. D39, a serotype 2 strain, was included as a negative control. The PCR primers for *mgrA* (SP1800F/SP1880R) are predicted to be specific for the *mgrA* gene, *Spr1622*, in D39 and should not amplify a product from *Spr1404*. All strains were positive by dot-blot and PCR for the presence of *mgrA* and *rpoB* but the islet genes were only present in strains of serotype 4, 9V and 6B (Figure 5.7).

MgrA or homolog is found in *S. pneumoniae* strains that do not possess the RlrA islet yet variation in expression of islet genes was the only consistent difference at all time points between the *mgrA* over-expressing and the *mgrA* deletion serotype 4 strains, as determined by the microarray experiments described above. It seems unlikely that the primary role of MgrA is to repress islet gene expression. It is more likely that MgrA has additional targets *in vivo* that were not detected by our microarray screen using RNA harvested from *in vitro*-grown bacteria and/or that it has alternative targets in non-serotype 4 *S. pneumoniae* strains.

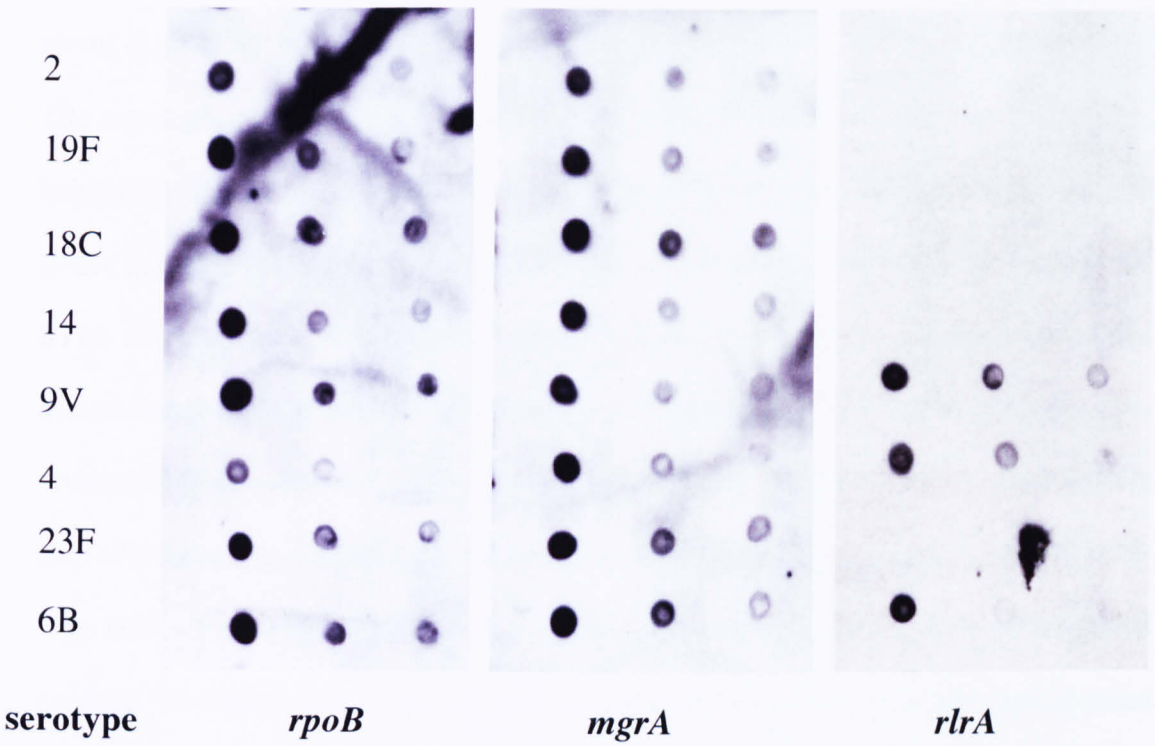


Figure 5.7 Presence of *mgrA* and *rlrA* genes in clinical strains of serotypes present in the pneumococcal conjugate vaccine, Prevnar.
DNA probes complementary to *mgrA* and *rlrA* were used to probe dot blots of chromosomal DNA from clinical isolates of different serotypes. Each DNA was spotted in serial dilutions across the membrane. An *rpoB* probe was used as a positive control. All probes were complementary to the TIGR4 gene sequence.

5.2.5 MgrA is a virulence factor in serotype 2 *S. pneumoniae*

The results of three STM screens of *S. pneumoniae* have been published to date. One was performed in a serotype 19F strain (Polissi *et al.*, 1998), a second in a serotype 3 strain (Lau *et al.*, 2001) and the most extensive in TIGR4 (Hava and Camilli, 2002). The *mgrA* gene is known to be present in all three serotypes but was only reported as being identified as a virulence factor in the TIGR4 STM screen. One possibility is that it is not required for disease in serotype 3 and 19F in the infection models used in the STM screens or that it was not identified in these screens as these were not comprehensive. To determine whether MgrA regulates any virulence factors in a non-serotype 4 *S. pneumoniae* strain a D39 strain (serotype 2) with an insertion deletion in *spr1622/mgrA* was constructed. This was assessed in competition assays against wild-type D39. Two models of infection were tested: lung infection and nasopharyngeal carriage. To construct serotype 2 strains with a deletion-insertion of a Spc^r gene in place of *mgrA* PCR amplification of the mutated region in the serotype 4 strain, AC1500, was performed using primer pairs 180AF1/180SRX and the purified amplicon transformed into *S. pneumoniae* D39. The double recombination event was selected for by plating on Spc. Confirmation of disruption of *spr1622* was confirmed by PCR and DNA sequencing.

The serotype 2 *mgrA* mutant, CH207, was attenuated in the pneumonia model (Figure 5.8) and the level of attenuation was greater than that caused by the corresponding mutation in TIGR4. It was not attenuated in a murine model of nasopharyngeal carriage. These data are consistent with MgrA being involved with regulation of factors in D39 that are important for invasive disease at a mucosal site but not for

nasopharyngeal carriage. As D39 does not contain the *rlrA* pathogenicity islet the MgrA regulator presumably is regulating other virulence factors in this strain.

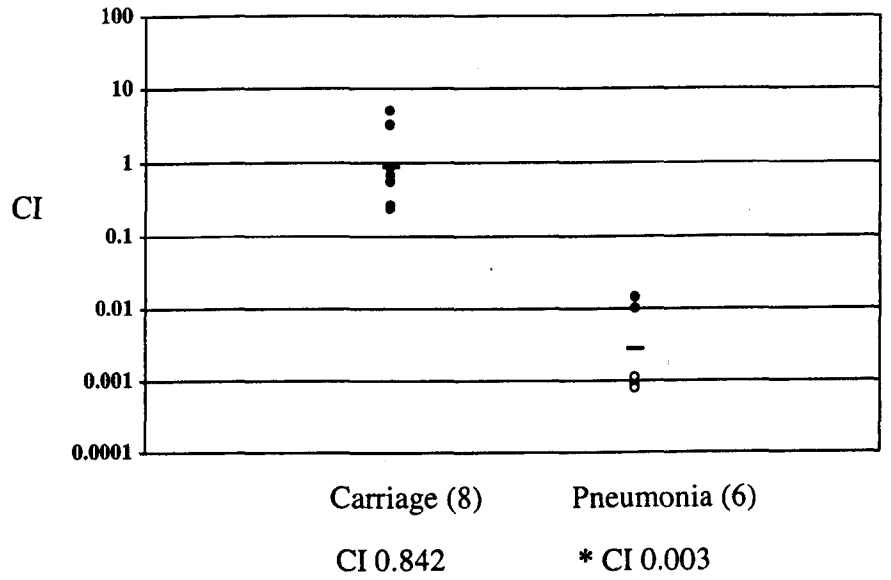


Figure 5.8 Competition assays D39 versus D39Δ*mgrA* (CH207)
Competition assays were performed using murine models of nasopharyngeal carriage and pneumonia. The *in vivo* competition index (CI) was calculated; each circle represents the CI derived from a single mouse in each set of competitions and the number of mice infected in each experiment is shown in parentheses. A CI of less than 1 indicates a virulence defect. Open circles indicate that no mutant bacteria were recovered from that animal and a value of 1 was substituted in the numerator when calculating the CI. The geometric mean of the CIs for all mice in a set of competitions is shown as a short solid line. The CI for pneumonia was statistically significant (*$p<0.05$).

Whilst performing these experiments it was noted that the inoculum required to generate sufficient bacterial numbers to determine a CI in competition assays in the murine model of pneumonia with either the TIGR4 or D39 *mgrA* mutants was 100-200 fold higher than is usually required for strains with mutations in other regulator-

encoding genes. This was repeatable over infections performed on a total of 6 separate occasions. To investigate this further 6 mice were infected with either 2 x 10^7 wild-type D39 bacteria, a similar dose of CH207 (D39 *mgrA* deletion-insertion strain) or a mixed inoculum comprising of 2 x 10^7 CFU of wild-type D39 and 2 x 10^7 CFU of CH207 (ie a total inoculum of 4 x 10^7 CFU). The input volume was adjusted to be the same for each mouse (40 μl). The infection was allowed to proceed for 44 hours in all but two mice, which were euthanised at 24 hours as they reached a sickness severity score requiring humane killing. Mice were scored for severity of illness, macroscopic lung abnormality at the time of lung dissection and the CFU/gram of tissue was recorded.

The experiment was repeated in duplicate on a separate day. The results given are from one experiment but are representative of both experiments (Table 5.2 and Figure 5.9).

As expected the mice infected with the attenuated CH207 strain alone showed no visible signs of infection and significantly less bacteria were recovered from the lungs compared to those infected with wild-type bacteria. Also, fewer mice, that were infected with the wild-type bacteria in combination with CH207, showed visible signs of infection, and significantly less bacteria were recovered from the lungs of these mice than the mice infected with wild-type alone (Table 5.2 and Figure 5.9). Therefore it appears that the *mrgA* mutant has the ability to restrict the growth of the wild-type bacteria *in vivo* and alter the course of the disease. Whether this is due to an alteration in host response to *S. pneumoniae* or due to an inherent change in the bacteria themselves is not clear. It is possible, for example, that they compete for sites needed to produce an infection such as a host cell-surface receptor.

Table 5.2 Affect of single and double infections: D39 and D39Δ*mgrA* (CH207)

Infective strain and Mouse number	State of health at 44 hours[a]	Macroscopic appearance of lung	CFU/gram lung
D39			
1	Sick (euthanised at 24 hours)	Red and friable	9.5×10^8
2	Sick (euthanised at 24 hours)	Red and friable	1.18×10^7
3	Sick	Mottled/lobar changes	2.45×10^7
4	Sick	Mottled/ lobar changes	2.59×10^3
5	Sick	Mottled/ lobar changes	7.83×10^3
6	Sick	Mottled/ lobar changes	1.03×10^4
CH207 (Δ*mgrA*)			
1	Well	Pink	0
2	Well	Pink	0
3	Well	Pink	0
4	Well	Pink	1.3×10^3
5	Well	Pink	1.76×10^3
6	Well	Pink	1.0×10^4
Mixed infection (D39 and CH207)			
1	Well	Pink	1
2	Well	Pink	4.5×10^2
3	Well	Pink	5.6×10^3
4	Well	Pink	7.4×10^3
5	Sick +/-	Mottled +/-	1.7×10^4
6	Sick +/-	Mottled +/-	4.6×10^4

[a] State of health at 44 hours except for D39 1 and 2 that were euthansed at 24 hours

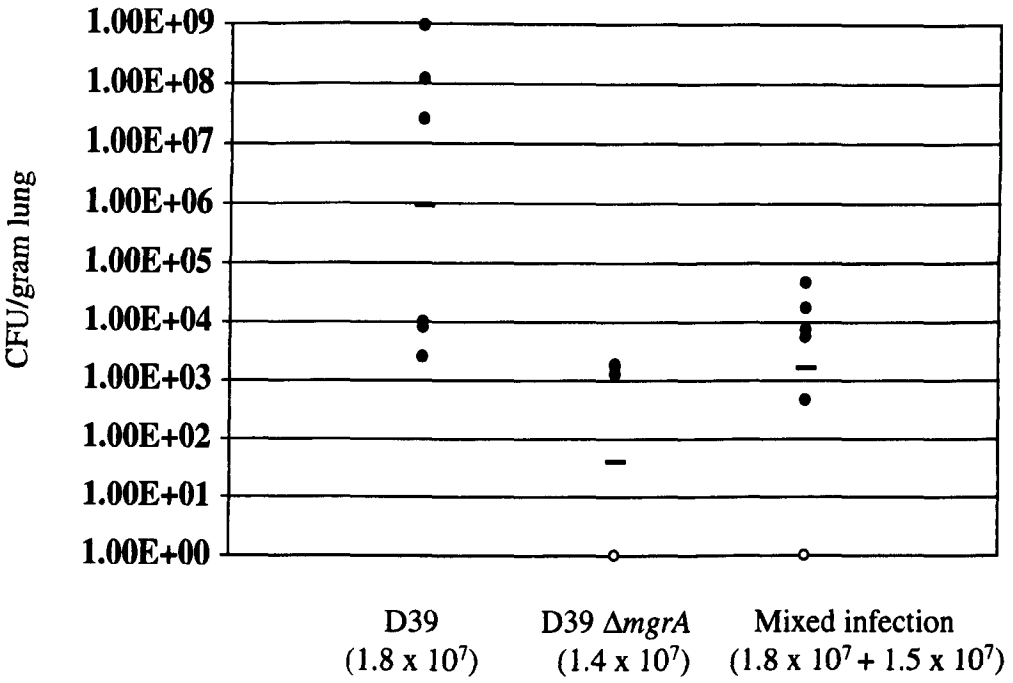


Figure 5.9 Murine infections using a pneumonia model: D39 and CH207 (D39Δ*mgrA*) single and mixed infection

Mice were inoculated intranasally with either D39 alone, CH207 (D39 Δ*mgrA*) alone or both strains. The input doses were as indicated in parentheses. Six mice were infected in each group, euthanased at 44 hours and bacteria harvested from the lung. Each circle represents the CFU/gram of lung tissue for a single mouse. Open circles indicate that no bacteria were recovered from that animal. The geometric mean is shown as a short solid line.

5.2.6 Are the genes of the *rlrA* pathogenicity islet the only targets of MrgA in serotype 4?

The work described in this chapter confirms that MgrA is a virulence factor in TIGR4 that represses the expression of genes within the *rlrA* pathogenicity islet. It is also a virulence factor in a second serotype, serotype 2. As the *rlrA* islet is not present on the genome in strains of this serotype, MgrA must be controlling the expression of factors other than the *rlrA* islet genes in serotype 2 strains. I wished to investigate whether MgrA is involved in controlling the expression of other factors in TIGR4 aside from the islet genes. It is quite possible that the *in vitro* growth conditions used to grow cells to harvest RNA for use on the microarrays were not appropriate for full activation of MgrA and hence not all targets were identified. The regulator may have a different state of activation *in vivo* and consequently regulate other factors *in vivo*.

I decided to approach this from two different angles; Firstly, I wished to determine what environmental signals MgrA might be responding to in order to develop better *in vitro* conditions for MgrA activation with a view to repeating microarray studies. Secondly, I wished to compare single and double *mgrA* and *rlrA* mutants in competition assays with TIGR4 strains in mice to assess whether the effects of the two mutations were additive or not.

5.2.7 Investigation of signals required for MgrA and RlrA activation

Mga and RofA in GAS control the expression of virulence factors in response to changing environmental conditions. Mga is activated by growth in elevated carbon-dioxide (Caparon *et al.*, 1992; McIver *et al.*, 1995). RofA is affected by superoxide levels and oxygen partial pressure (Fogg and Caparon, 1997; Kreikemeyer *et al.*, 2003).

Nra, however, does not appear to be regulated by similar conditions to RofA (Kreikemeyer *et al.*, 2003). There is some evidence that the ambient oxygen concentration is an important factor in the ability of *S. pneumoniae* to regulate the characteristics of its cell surface. Oxygen levels affect the opaque and transparent phase variants differently. There is an increased production of capsular polysacharride (CPS) in opaque variants in conditions of reduced oxygen, whereas synthesis of CPS in transparent variants remains comparatively low under atmospheric or reduced oxygen conditions. To determine if oxygen and/or carbon dioxide levels are involved in the activation status of MgrA and RlrA in *S. pneumoniae* a *lacZ* reporter gene was used to examine the transcription of genes, *rlrA* and *srtD* in different environments. These genes were selected on the basis that they are known to encode virulence factors and are targets for MrgA and RlrA. There were no other known targets outside the islet for either regulator. Several strains were constructed: two with transcriptional fusions of the *lacZ* gene to *rlrA* or *srtD* (CH210 and CH214 respectively) and strains with single or double deletions in *mgrA* and *rlrA* in the *lacZ* fusion backgrounds (CH211, CH212 and CH213) (Table 2.3).

The affect of growth phase on transcription of *rlrA* and *srtD* was assessed. CH214, CH210 and wild-type, AC353, were grown in THY broth and aliquots taken at the same 4 time points in the growth curve as had been used when harvesting RNA (early exponential, mid exponential, late exponential and early stationary phase). β-galactosidase activity was measured from the lysed cell pellet (Figure 5.10).

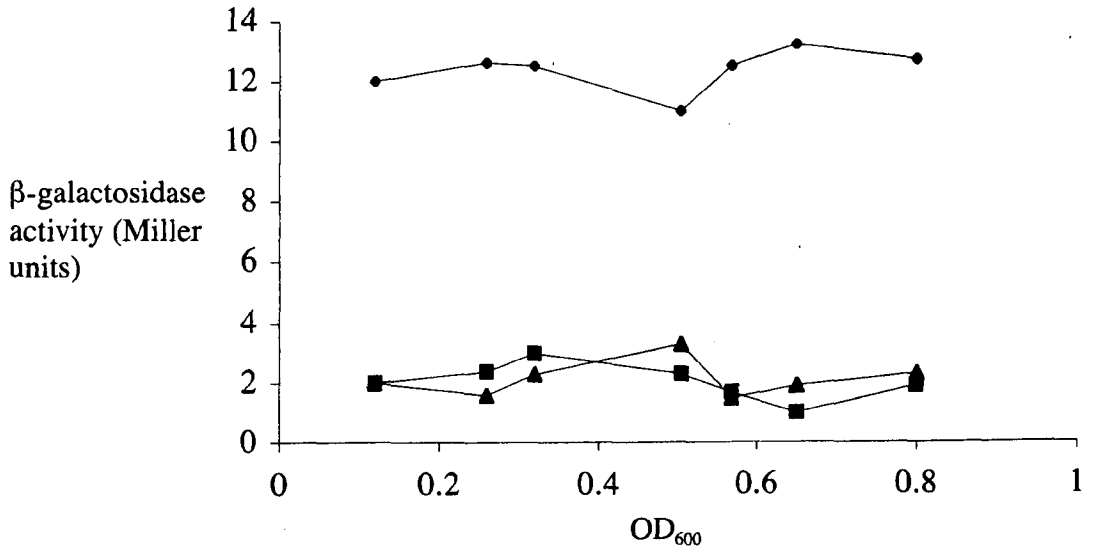


Figure 5.10 β-galactosidase activity of *srtD::lacZ* and *rlrA::lacZ* fusion strains
CH210 (◆, *srtD::lacZ*), CH214 (▲, *rlrA::lacZ*) and AC353 (■, wild-type) were grown in THY broth and aliquots taken at the OD_{600} indicated. β-galactosidase activity was measured from the lysed cell pellets.

The level of activity seen in the *srtD::lacZ* strain was greater than that seen in the *rlrA::lacZ* strain and activity in both was constant over time. As there was no change in β-galactosidase with growth all cells in subsequent experiments were grown to mid-exponential phase. Experiments were repeated in triplicate.

The effect of altering the atmosphere was examined. β-galactosidase activity was measured in cells grown to mid-exponential phase in normal atmospheric conditions (0.03% carbon dioxide, 20% oxygen), anaerobically (10% carbon dioxide, 0.03% oxygen) or in 10% carbon dioxide, 20% oxygen. In the *srtD::lacZ* strain β-galactosidase activity was highest in cells grown in 10% carbon dioxide and 20% oxygen. Activity was repressed by reducing either the oxygen level or the carbon

dioxide level (Figure 5.11) Deletion of *mgrA* in this background abolished the difference in activity seen in varying atmospheres indicating that MgrA was required for this effect. RlrA was absolutely required for *srtD* transcription as inactivation of *rlrA* in the *srtD::lacZ* background reduced the β-galactosidase activity to the background levels seen in wild-type cells (Figure 5.11). Additional deletion of *mgrA* in the *srtD::lacZ* Δ*rlrA* strain had no effect on β-galactosidase activity consistent with the idea that MgrA works to repress transcription of *srtD* by repressing *rlrA* transcription.

In the *rlrA::lacZ* strain β-galactosidase activity was similar to that seen in wild-type cells (Figure 5.10). The level was increased in the *mgrA* deletion strain by 2-3 fold. There was no difference in the β-galactosidase activity measured in different atmospheric conditions indicating that RlrA was not affected by changing carbon dioxide or oxygen levels (Figure 5.11). However the level of activity of the fusions was so low that it is possible that this is not an adequate method to test this hypothesis. The data is consistent with the activation status of MgrA being affected by oxygen and/or carbon dioxide levels. *In vitro* growth conditions that resulted in improved MgrA activation compared to those used for the microarray screen could not be determined.

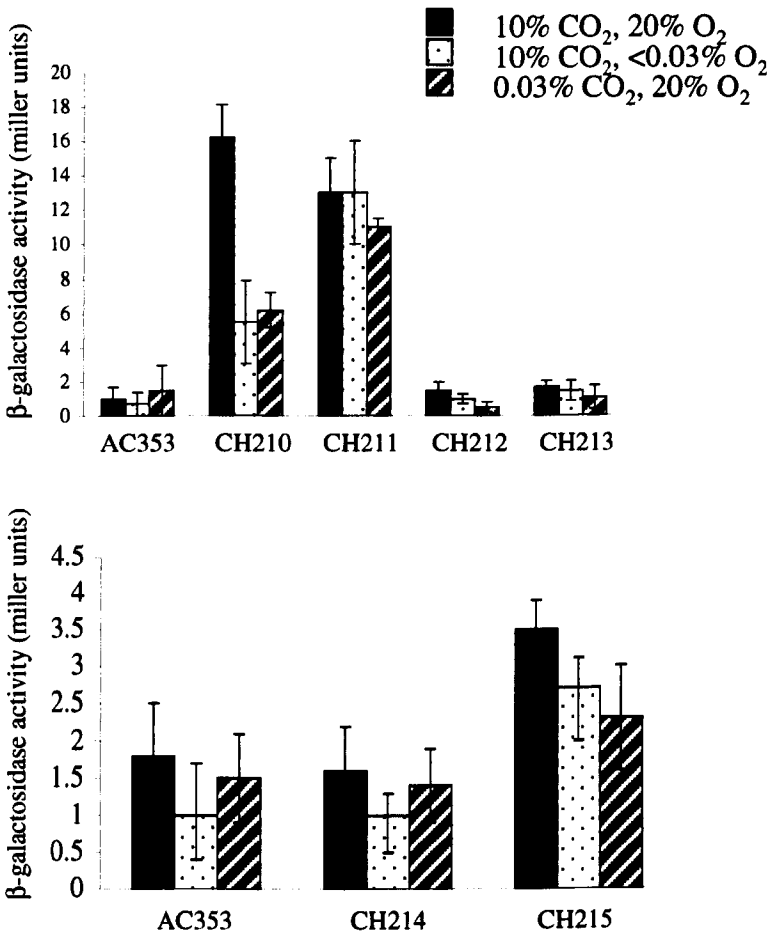


Figure 5.11 β-galactosidase activity of (A) *srtD::lacZ* fusion strains (B) *rlrA::lacZ* fusion strains in different atmospheres.

(A) AC353 (wild-type), CH210 *(srtD::lacZ)*, CH211 (*srtD::lacZ ΔmgrA*), CH212 (*srtD::lacZ rlrA::magellan2*) and CH213 (*srtD::lacZ rlrA::magellan2 ΔmgrA*)

(B) AC353 (wild-type), CH214 *(rlrA::lacZ)* and CH215 (*rlrA::lacZ ΔmgrA*)

All *S. pneumoniae* strains were grown in THY broth to mid exponential phase. β-galactosidase activity was measured from the lysed cell pellets. Each strain was assessed in different growth atmospheres as indicated. Data for strains grown in 5%CO_2, 20% O_2 is not shown but similar to results for growth in 10%CO_2, 20% O_2.

5.2.8 MgrA and RlrA act in the same pathway

Whether two virulence genes interact as part of the same function or act independently can be investigated using a test referred to as a cancelled out index or COI (Beuzon and Holden, 2001). The COI is calculated by competition of each single mutant against the double mutant. If the two genes contribute equally to the same function, the double mutant should not be more attenuated than the single mutant. Explained another way; the CI of wild-type versus the single mutant should be less than one, whereas the COIs should not be different from one. If the COI equals one it is possible that the action of one gene product is dependent on the gene product of the second gene i.e. they work in the same pathway.

If the *rlrA* islet genes are the only targets for MgrA in TIGR4 then one would predict that the Δ*mgrA rlrA::magellan2* mutant (CH209) would have a similar *in vivo* phenotype to the *rlrA::magellan2* mutant strain (AC1213). To determine if this is the case a double regulator mutant was constructed. This strain, CH209 was tested in competition assays with AC1213 for pneumonia and nasopharyngeal carriage (COI). The COI were not significantly different from 1 for either type of infection (Figure 5.12). There was no additional attenuation over and above that seen in the *rlrA* mutant with the addition of a second *mgrA* mutation, consistent with the idea that both regulators act within the same pathway. These results are consistent with the hypothesis that MgrA does not regulate any other factors required for causing lung infection or nasopharyngeal carriage in a murine model in TIGR4 other than those in the *rlrA* islet.

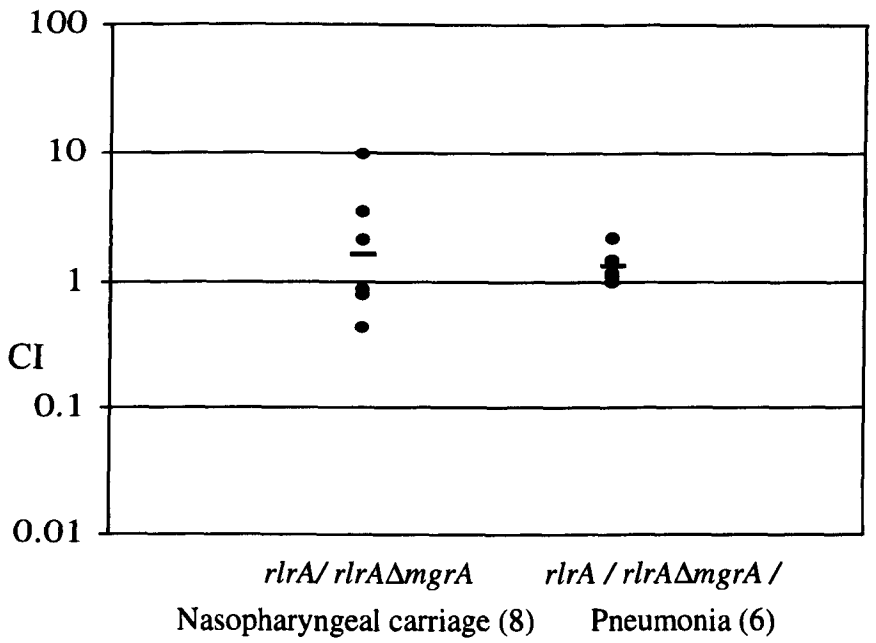


Figure 5.12 COIs of *rlrA::magellan2* mutant versus *ΔmgrA rlrA::magellan2* double mutant

Competition assays/COIs were performed using murine models of carriage and pneumonia. The COI was calculated; each circle represents the result derived from a single mouse in each set of competitions and the number of mice infected in each experiment is shown in parentheses. The geometric mean for all mice in a set of competitions is shown as a short solid line. No COI was statistically different from 1.

5.3 SUMMARY

In this chapter I have described work to identify potential targets for the transcriptional regulator MgrA. With the aid of cDNA microarrays the genes of the *rlrA* pathogenicity islet were identified as being repressed by MgrA at all time points across the growth curve. This was confirmed by RPAs probing separately prepared RNA for transcript of one gene from each of the four operons within the islet. Further RPAs determined that, unlike their counterparts in GAS, RlrA and MgrA did not appear to show reciprocal regulation. Genes in a second putative operon, the ZmpB operon, were identified by the microarray screen as possibly being transcriptionally activated by MgrA at late time points in the growth curve. This however was not confirmed by RPA.

Published data on the *S. pneumoniae* genome in two serotypes (types 2 and 4) shows that the *rlrA* islet is not present in serotype 2 but *mgrA* is. Work described here extends this finding further with the use of dot blots probing for genes in total chromosomal DNA from 8 different clinical isolates representative of all the serotypes included in the pneumococcal conjugate vaccine. The *mgrA* gene was present in all strains but the islet was only detected in 3 serotypes, 4, 9V and 6B. Competition assays showed MgrA to be a virulence factor in the serotype 2 D39 strain that does not possess the *rlrA* islet. This indicates that MgrA does control other factors involved in causing disease in at least one non-serotype 4 strain. In contrast there was no evidence that MgrA controls factors other than the islet genes in serotype 4 as a serotype 4 Δ*mgrA rlrA::magellan2* double mutant showed no greater attenuation than the *rlrA::magellan2* single mutant as assessed by competition assays in mice.

Studies using *lacZ* fusion strains showed that oxygen and carbon dioxide levels affected the level of transcription of the *rlrA* and *SrtD* genes of the *rlrA* islet. This required the presence of *mgrA* indicating that MgrA was involved, at least in part, in this regulation.

Acknowledgements

I thank Stanley Falkow for supporting my stay at his laboratory, Stanford University, California, USA to enable me to perform the microarray experiments and Elizabeth Joyce for technical assistance and advice during this time. Elizabeth Joyce also performed the duplicate microarray experiments.

CHAPTER 6 FUNCTION AND DISTRIBUTION OF THE *RLRA* PATHOGENICITY ISLET

6.1 INTRODUCTION

In the previous two chapters I have shown that two transcriptional regulators RlrA and MgrA are required for virulence in a murine model of pneumonia and nasopharyngeal carriage. Both are involved in regulation of transcription of the genes on the *rlrA* pathogenicity islet. It has been demonstrated that islet-encoded transcriptional regulator, RlrA, is a positive regulator of each of the genes in the islet and that MgrA, encoded by a gene outside the islet acts as a repressor of islet gene transcription. No other targets have been identified for either regulator. Interestingly the *rlrA* islet is not present in the chromosome of all clinical isolates of *S. pneumoniae* and it is likely that the islet was acquired by a serotype 4 strain by horizontal transfer from another *S. pneumoniae* strain or another Gram-positive species. This idea is further supported by the fact that it is flanked by two IS1167 insertion sequences allowing integration of the element into the genome to have occurred by homologous recombination or a transposition event. Gram-positive bacteria in general are known to contain mobile elements (Butaye *et al.*, 2003; Gentry-Weeks *et al.*, 2002; Lindsay *et al.*, 2001). The *S. pneumoniae* genome itself is rich in insertion sequences (ISs) which make up ~5% of the TIGR4 chromosome, corresponding to >3.5% (84 out of 2236) of the genes in this organism (Tettelin *et al.*, 2001). However the majority of these insertion sequence elements appear to be non-functional due to additional insertions, deletions and point mutations resulting in frames shifting or degenerate transposase genes. It conceivable that the DNA transformation system in *S. pneumoniae* may allow conversion of disrupted insertion sequence elements to functional ones by homologous recombination. The same system does allow

uptake of extra-cellular DNA and homologous recombination into or around IS sites on the chromosome. *S. pneumoniae,* therefore, could change its genetic make-up from either transposition events or by homologous recombination and in this way possibly gain new virulence factors. If the *rlrA* islet is transferable between strains of *S. pneumoniae,* its transfer would provide a means for the acquisition of new virulence factors. Data presented in chapter five shows that the element has not only inserted into the genome but that the genes have become integrated into at least one regulatory network within *S. pneumoniae* serotype 4. Based on the islet's organization, internal regulation and its presence in only a subset of strains, I hypothesised that the islet element contains the information necessary to function in any strain of *S. pneumoniae* that acquires it without the need for additional genetic alterations and that the surface adhesins would be expressed and functional.

Work presented in this chapter assigns an *in vitro* phenotype associated with the *rlrA* islet, demonstrates that transfer of the *rlrA* element to a strain lacking the islet can result in a gain of function and investigates the occurrence of *in vivo* islet transfer in a murine model of nasopharyngeal carriage. The distribution of the islet in clinical strains was also investigated further to look for any correlation between virulence and islet presence.

6.2 RESULTS

6.2.1 The *rlrA* locus is involved in binding a human lung epithelial cell line *in vitro*

Three genes on the islet, *rlrA, rrgA* and *srtD*, are required for lung infection and/or nasopharyngeal carriage in mice (Hava and Camilli, 2002; Hava *et al.*, 2003a). The

rrgA gene is predicted to encode for an MSCRAMM suggesting that it has a role in binding to extracellular matrix molecules. However, it does not possess classical fibronectin- or collagen- binding regions as characterized in other Gram-positive organisms (Cue *et al.*, 2001; Jaffe *et al.*, 1996). It does, however, possess other domains found in MSCRAMMS such as a Cna protein B-type domain and von Willebrand factor (vWF) type A domain (Figure 6.1). The Cna domain is found in *Staphylococcus aureus* collagen-binding surface proteins and is thought to form a stalk in the protein that presents the actual ligand binding domain away from the bacterial cell surface (Snodgrass *et al.*, 1999). The vWF domain is found in various proteins of which a common feature is the involvement in multi-protein complexes. Proteins that possess the vWF type A domain participate in events such as cell adhesion, migration and signal transduction (Doliana *et al.*, 1993). The sortase encoded for by the *srtD* islet gene selectively recognizes and attaches RrgA to the cell wall (Hava and Camilli, personal communications). The other two sortases encoded by islet genes recognize the two other surface proteins, RrgB and RrgC, and do not act on RrgA (Hava and Camilli, personal communications). Given the requirement for SrtD on RrgA cell wall anchoring it follows that *rrgA* and *srtD* mutants have similar phenotypes.

The human lung epithelial cell line, A549, has been used extensively as a model of the *S. pneumoniae* host mucosal cell interaction (Cundell *et al.*, 1995a). These cells were therefore chosen for analysis of the involvement of RlrA, RrgA and SrtD in bacterial adhesion to host mucosal surfaces. The ability of wild-type AC353 and isogenic strains (AC1213, AC1215, AC1214) possessing transposon-insertions in each of the three genes (*rlrA*, *rrgA* and *srtD* respectively) to adhere to A549 cells, was analysed. The transposon insertions were not believed to have polar effects on other genes in the islet

1 MLNRETHMKKVRKIFQKAVAGLCCISQLTAFSSI
35 VALAETPETSPAIGKVVIKETGE GGALLGDAVF
69 ELKNNTDGTTVSQRTEAQTGEAIFSNIKPGTYTL
103 TEAQPPVGYKPSTKQWTVEVEKNGRTTVQGEQV
136 ENREEALSDQYPQTGTYPDVQTPYQIIKVDGSEK
170 NGQHKALNPNPYERVIPEGTLSKRIYQVNNLDDN
204 QYGIELTVSGKTVYEQKDKSVPLDVVILLDNSNS
237 MSNIRNKNARRAERAGEATRSLIDKITSDSENRV
271 ALVTYASTIFDGTEFTVEKGVADKNGKRLNDSLF
306 WNYDQTSFTTNTKDYSYLKLTNDKNDIVELKNK
339 VPTEAEDHDGNRLMYQFGATFTQKALMKADEIL
373 TQQARQNSQKVIFHITDGVPTMSYPINFNHATFA
407 PSYQNQLNAFFSKSPNKDGILLSDFITQATSGEH
441 TIVRGDGQSYQMFTDKTVYEKGAPAAFPVKPEK
474 YSEMKAAGYAVIGDPINGGYIWLNWRESILAYP
507 FNSNTAKITNHGDPTRWYYNGNIAPDGYDVFTV
540 GIGINGDPGTDEATATSFMQSISSKPENYTNVTD
575 TTKILEQLNRYFHTIVTEKKSIENGTITDPMGELI
610 DLQLGTDGRFDPADYTLTANDGSRLENGQAVGG
643 PQNDGGLLKNAKVLYDTTEKRIRVTGLYLGTDE
676 KVTLTYNVRLNDEFVSNKFYDTNGRTTLHPKEV
709 EQNTVRDFPIPKIRDVRKYPEITISKEKKLGDIEF
744 IKVNKNDKKPLRGAVFSLQKQHPDYPDIYGAID
777 QNGTYQNVRTGEDGKLTFKNLSDGKYRLFENSE
810 PAGYKPVQNKPIVAFQIVNGEVRDVTSIVPQDIP
844 AGYEFTNDKHYITNEPIPPKREYPRTGGIGMLPF
877 YLIGCMMMGGVLLYTRKHP

Figure 6.1 RrgA sequence
Amino acid sequence of RrgA. CnaB domains are highlighted in red and the vWF type A domain in blue. The sortase recognition YPXTG motif is highlighted in bold and is followed by a stretch of hydrophobic residues (yellow) and a charged tail.

as the transposon was either in a singly transcribed gene (*rlrA* and *rrgA*) or the last gene of an operon (*srtBCD*). (Figure 7.1). Adhesion assays were performed as described in section 2.5.1. A549 cells were grown to ~ 90-95% confluence. *In vitro* grown bacteria were added to the monolayer at an MOI of 10:1. An average of 0.3% of the input wild-type cells adhered to A549 cells. Appropriate negative controls were included to determine if bacteria adhered to the plastic surface of the well or the glass coverslip incubated with serum alone. No bacteria were recovered from PBS-washed wells that had not been seeded with A549 cells, but had been pre-incubated with culture medium, confirming that *S. pneumoniae* did not adhere to the wells or coverslips alone. All experiments were performed in quadruplicate and each experiment was replicated 5 times on different days. Transposon insertions in *rlrA* and *rrgA* reduced binding of *S. pneumoniae* to the epithelial cell layer by 5- to 10-fold (Figure 6.2), suggesting that *rrgA* is important for interactions between TIGR4 and lung epithelial cells. The strain with a transposon insertion in *srtD* did not produce consistent results. Unlike the other strains, there was a high degree of inter-experimental variability for this strain. On three occasions *srtD* mutant binding was at wild-type level and twice binding was reduced to 20-50% of wild-type.

To ascertain if the islet has a similar role in another *S. pneumoniae* serotype that possesses the element, the binding of a wild-type serotype 9V strain (AC1365) was compared to isogenic *rlrA* and *rrgA* mutants (CH185 and CH184 respectively). An average of 0.1% of the input wild-type serotype 9V cells adhered to A549 cells. The overall level of binding was decreased in the 9V strain compared to the serotype 4 strain, binding was similarly dependent on the presence of the RlrA regulator and RrgA.

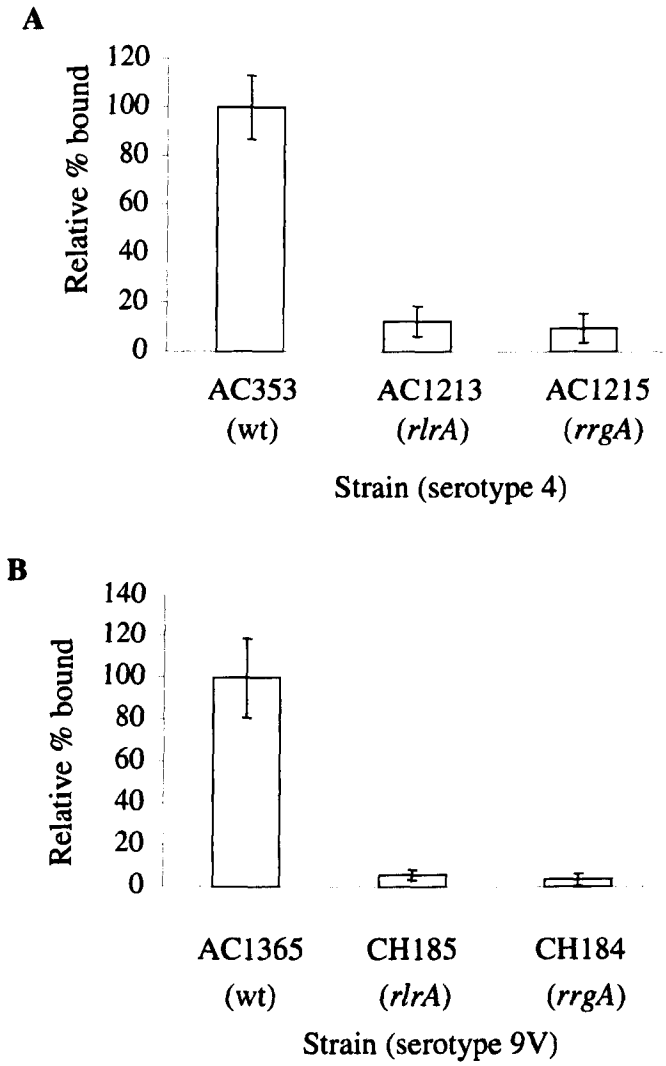


Figure 6.2 Adherence of serotype 4 and 9V *S. pneumoniae* strains to A549 lung epithelial cells

A549 monolayers were inoculated with (**A**) Serotype 4 strains AC353 (wild-type), AC1213 (*rlrA::magellan2*) and AC1215 (*rrgA::magellan5*) or (**B**) Serotype 9V strains AC1365 (wild-type), CH185 (*rlrA::magellan2*) and CH184 (*rrgA::magellan5*) *S. pneumoniae* strains. Mid-exponential growth phase bacterial cells ($1x10^6$ CFU) were added to A549 cells at a multiplicity of infection of 10:1 and incubated for 40 minutes. The percentages of bacteria that adhered to the monolayers after three washes with PBS were determined. Triton/EDTA was added to each well to detach the cells and liberate bacteria, which were then enumerated by plating lysates on blood agar plates. The adherence of each strain is expressed relative to the adherence of the wild-type (serotype 4; AC353 or serotype 9V; AC1365). Representative results from one of three independent experiments are shown. The assay was performed in quadruplicate in each experiment and the error bars represent the standard deviations.

Transposon insertions into *rlrA* and *rrgA* reduced the binding of *S. pneumoniae* to the epithelial cell layer by 10 to 20-fold of the wild-type level (Figure 6.2).

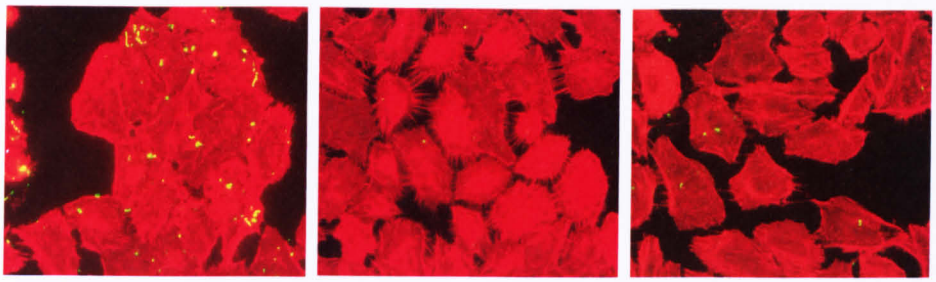
Immunofluorescence microscopy was used as an alternative method to assess adherence of wild-type, *rrgA* and *rlrA* mutant strains to the monolayer. A549 monolayers were grown on glass coverslips in 24-well tissue culture plates. *In vitro* grown serotype 4 bacteria suspended in Ham's F12K tissue culture medium plus 10% fetal bovine serum were added to the monolayer at an MOI of 10:1. The plates were spun at 1000 rpm for 2 minutes, to increase initial contact between bacteria and cells, then incubated for 40 minutes at 37°C and 6% CO_2 after which time the culture fluid was removed from each well, the monolayers were washed 3 times with PBS (pH 7.4) to remove non-adherent bacteria and fixed in 3% PFA in PBS. Bacteria were labeled using an anti-capsular *S. pneumoniae* primary antibody and a Cyanine-2 (Cy2)–conjugated donkey anti-rabbit secondary antibody. Epithelial cells were permeabilised with 0.1% saponin in PBS and labeled with Rhodamine-conjugated phalloidin to detect F-actin.

A control experiment was carried out to confirm that the centrifugal sedimentation of bacteria affected all strains similarly. Coverslips were removed after the centrifugation step, prior to incubation and the cells present on the coverslips were directly fixed and labeled. The number of bacteria per epithelial cell was counted. Adjusted was made for actual input number. Over 200 epithelial cells were examined. There was no difference between the number of wild-type, *rlrA* and *rrgA* mutant strains associated with host cells after centrifugation and prior to the 40 minute incubation (data not shown).

Over 200 epithelial cells were examined for *S. pneumoniae* binding after the period of incubation and subsequent washes and fixing. Similar to the results seen by counting recovered CFU, binding was dramatically reduced by inactivation of either *rlrA* or *rrgA*. The level of reduction calculated by this method was greater than by counting recovered CFU. Binding of *S. pneumoniae* was reduced 20- to 40-fold in AC1213 (*rlrA* mutant) and AC1215 (*rrgA* mutant) respectively (Figure 6.3).

6.2.2 Binding to A549 cells is affected by MgrA

Deletion of *mgrA* increases the expression of all seven islet genes *in vitro*. However this *in vitro* regulatory effect is inconsistent with the *in vivo* phenotype displayed by the Δ*mgrA* mutant (Table 4.1). One might predict that an Δ*mgrA* mutant would have an opposite *in vivo* phenotype to an *rlrA* mutant. This is not the case. Both are less proficient at causing pneumonia or carriage than wild-type bacteria. To determine if the level of transcriptional de-repression of the islet genes seen in the *mgrA* deletion mutant corresponded to a measurable phenotypic effect on bacterial adherence, the ability of wild-type, *rlrA* mutants (AC1213 and AC1278) and *mgrA* mutants (AC1500 and AC1481) to adhere to A549 cells was analyzed. All strains were grown in THY 0.8% maltose prior to infection of the A549 monolayer as this media was required for maximal expression of *mgrA* or *rlrA* in the over-expressing mutants. The four mutant strains exhibited either increased or decreased adherence compared to wild-type depending on the nature of the mutation. The transposon insertion in *rlrA* reduced binding of *S. pneumoniae* to A549 cells to 10% of wild-type and conversely over-expression of *rlrA* increased binding to 160%. Over-expression of *mgrA* reduced binding to 30% and deletion of *mgrA* increased binding to 140% of wild-type (Figure 6.4).



AC353 (wild-type) AC1213 (*rlrA*) AC1214 (*rrgA*)

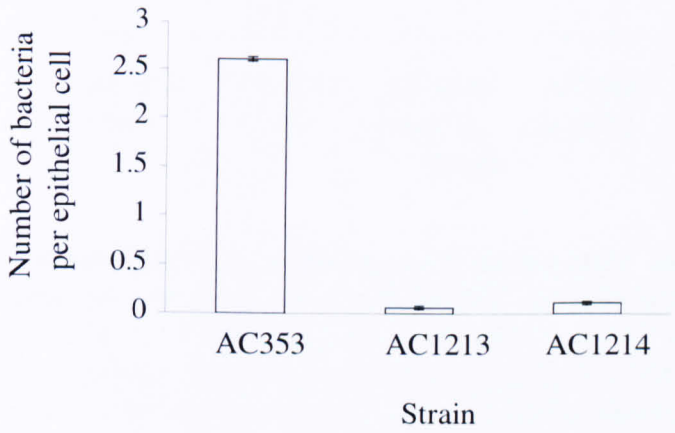


Figure 6.3 Immunofluoresence microscopy of serotype 4 *S. pneumoniae* strains adhering to A549 lung epithelial cells

A549 monolayers were inoculated with AC353, AC1213 (*rlrA:: magellan2*) and AC1214 (*rrgA:: magellan5*) *S. pneumoniae* strains. Mid-exponential growth phase bacterial cells ($1x10^6$ CFU) were added to A549 cells at a multiplicity of infection of 10 and incubated for 40 minutes. (**A**) Coverslips were fixed after three washes with PBS. Bacteria were labeled with anti-capsular antibody (green) and F-actin of epithelial cells were visualised by staining with Rhodamine-conjugated phalloidin after permeabilisation with 0.1% saponin (red). (**B**) The number of bacterial cells per epithelial cell of each strain is expressed. Representative results from one of three independent experiments are shown. The assay was performed in quadruplicate in each experiment and the error bars represent the standard deviations.

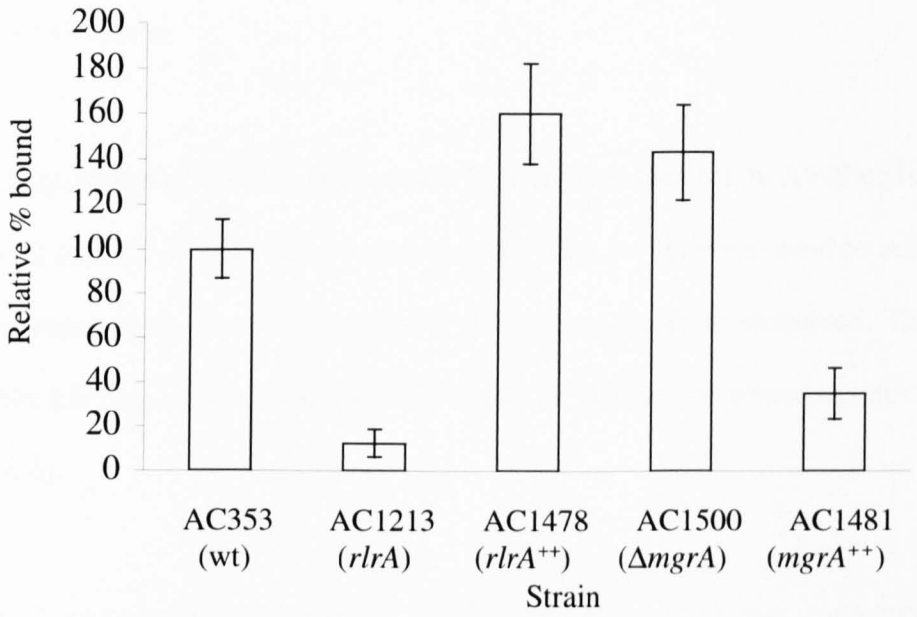


Figure 6.4 Role of MgrA in adherence of *S. pneumoniae* to A549 cells
A549 monolayers were inoculated with AC353, AC1213 (*rlrA:: magellan2)*, AC1478 (*rlrA*$^{++}$), AC1500 (*ΔmgrA*) or AC1481 (*mgrA*$^{++}$) *S. pneumoniae* strains. Mid-exponential growth phase bacterial cells ($1x10^6$ CFU) were added to A549 cells at a multiplicity of infection of 10 and incubated for 40 minutes. The percentages of bacteria that adhered to the monolayers after three washes with PBS were determined. Triton/EDTA was added to each well to detach the cells and liberate bacteria, which were then enumerated by plating lysates on blood agar plates. The adherence of each strain is expressed relative to the adherence of the wild-type. Representative results from one of three independent experiments are shown. The assay was performed in quadruplicate in each experiment and the error bars represent the standard deviations.

These data show that the alteration of the islet gene expression by MgrA seen *in vitro* is correlated with a change in adherence phenotype of the pneumococcal strain. However this *in vitro* phenotypic change does not appear to correlate with the *in vivo* phenotype of reduced virulence.

6.2.3 Acquisition of the *rlrA* islet results in increased binding to A549 cells

Serotype 2 strain D39 does not contain the *rlrA* islet. Its ability to bind to A549 cells using the same protocol as for serotype 4 strains was therefore examined. There was negligible binding as assessed by CFU count or immunofluorescence microscopy (Figure 6.6).

The region surrounding *rlrA* islet on the TIGR4 chromosome was compared to the corresponding region in the R6 (a derivative of D39) genome sequence. Identical regions to those flanking the *rlrA* islet and neighbouring IS elements in TIGR4 were found in R6. Corresponding genes on the two chromosomes are greater than 99% identical at the nucleotide level. Interestingly though, the R6 sequence does not contain an IS element at the *rlrA* islet locus. Instead there is a 686 bp sequence in R6 that is not present in TIGR4 (Figure 6.5). To determine if acquisition of the islet by a strain that did not possess it resulted in a change of phenotype the *rlrA* islet was transferred into D39 by transformation of competent D39 cells with chromosomal DNA from a serotype 4 strain that had a transposon insertion in the *rlrA* proximal IS element. The presence of the transposon allowed for antibiotic selection of transformants that gained the IS element. The presence of the islet in D39 was confirmed by PCR using chromosomal and islet-specific primers for each end of integration (Figure 6.5). PCR products could be detected only if integration of the whole islet had occurred at that specific site.

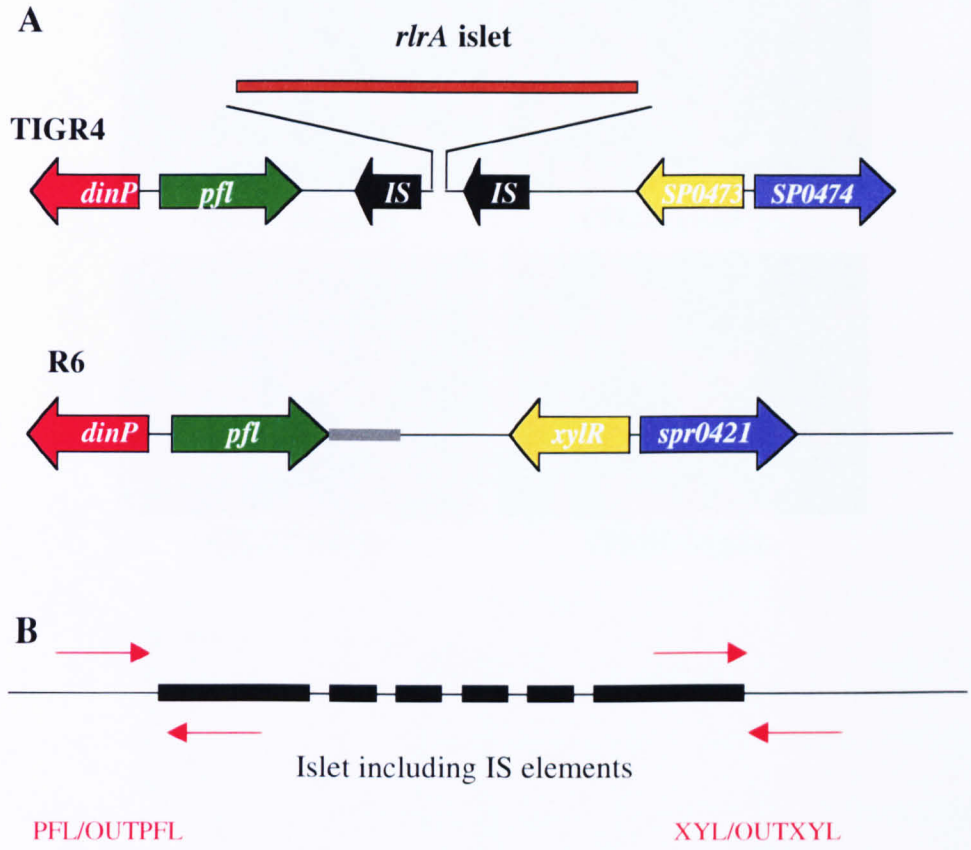


Figure 6.5 Islet integration site

(**A**) The site of *rlrA* islet integration in TIGR4 is shown compared to the comparable site in the R6 genome. Genes, indicated as arrows, of the same colour are identical at the nucleotide level. The grey line in R6 after *pfl* indicates a 686 bp sequence that is not present in the TIGR4 genome sequence. (**B**) PCR confirmation of the site of integration using primer sets (PFL/OUTPFL and XYL/OUTXYL, indicated as red lines) corresponding to each end of the islet were used to amplify the junction regions.

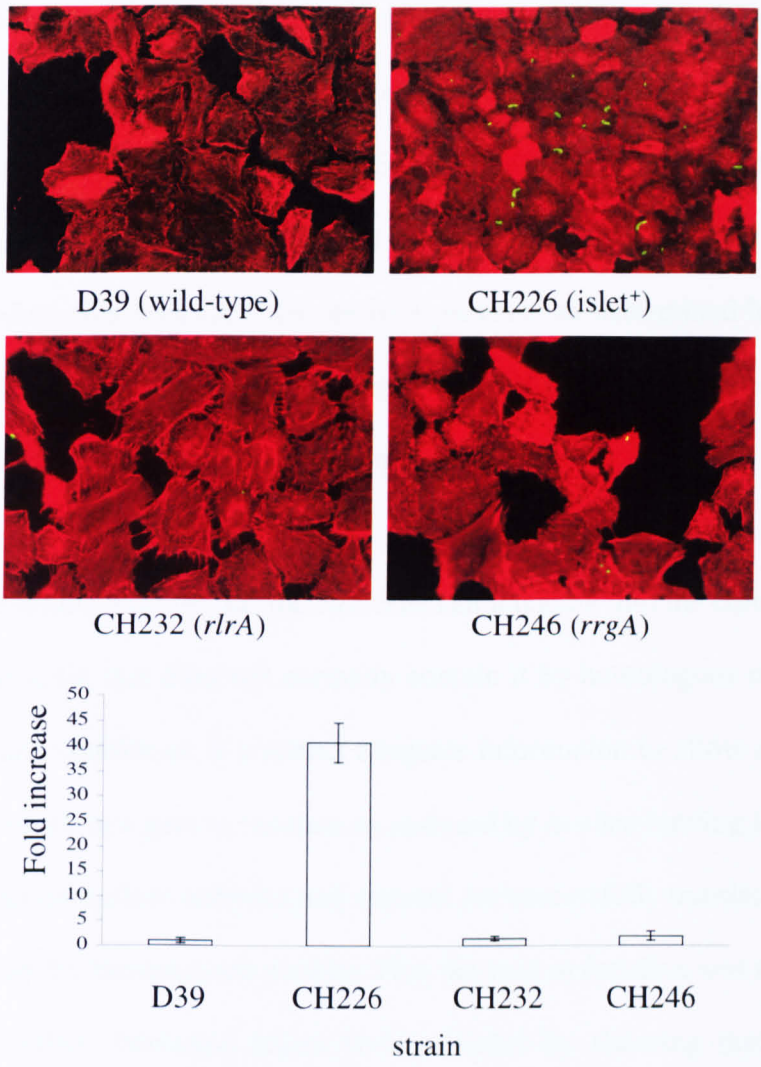


Figure 6.6 Immunofluoresence microscopy of serotype 2 *S. pneumoniae* strains adhering to A549 lung epithelial cells

A549 monolayers were inoculated with D39 (wild-type), a serotype 2 strain possessing the islet CH226 (*IS1167:: magellan5*), CH232 (*IS1167:: magellan5 rlrA::magellan2*) and CH246 (*rrgA::*magellan2) S. *pneumoniae* strains. Mid-exponential growth phase bacterial cells ($1x10^6$ CFU) were added to A549 cells at a multiplicity of infection of 10:1 and incubated for 40 minutes. (**A**) Coverslips were fixed after three washes with PBS. Bacteria were labeled with anti-capsular antibody (green) and F-actin in epithelial cells were visualised by staining with Rhodamine-conjugated phalloidin after permeabilisation with 0.1% saponin (red). Short chains of *S. pneumoniae* bacteria can be seen in green. (**B**) The number of bacterial cells per epithelial cell of each strain was calculated. The fold increase compared to wild-type D39 is reported. Representative results from one of three independent experiments are shown. The assay was performed in quadruplicate in each experiment and the error bars represent the standard deviations.

Confirmation of whole islet transfer was also obtained by ability to PCR amplify each of the seven genes from transformants. 20 transformants were screened by PCR and all showed successful integration of the entire islet. Two of these were assayed for binding to A549 cells. Results for only one (CH226) are shown but are representative of both strains. Acquisition of the islet resulted in measurable binding to A549 cells (Figure 6.6). The binding was dependent on the islet presence as determined by subsequent inactivation of the *rlrA* or *rrgA* gene (Figure 6.6). This shows that binding required intact islet genes and specifically *rlrA* and *rrgA*.

Together these results indicate that the *rlrA* islet can integrate into the chromosome of a *S. pneumoniae* strain that does not normally contain it by homologous recombination under appropriate conditions. It contains adequate information to allow acquisition of the element to result in a gain in function as assessed by *in vitro* binding to A549 cells. This implies that the surface adhesins and sortases are successfully translated, processed and expressed on the bacterial cell surface. That the gain in function was not due to the acquisition of other, unrelated genes was excluded by showing that subsequent inactivation of *rlrA* or *rrgA* results in loss of or reduction in binding.

To determine whether integration of the islet into the D39 genome was sufficient to allow integration into endogenous regulatory networks strains were constructed that possessed the islet but that also either had a deletion in *mgrA* (CH207) or over-expressed *mgrA* (CH216). Binding of these two strains to A549 cells was compared to that of wild-type D39. There was no conclusive evidence in these assays that the endogenous MgrA regulator affected transcription of the newly acquired islet genes as,

unlike in AC353 (Figure 6.4), inactivation or over-expression of *mgrA* in the serotype 2 background did not significantly affect binding to A549 cells (Figure 6.7).

6.2.4 Acquisition of the *rlrA* islet results in increased virulence

I wished to determine if islet acquisition also led to a change in virulence. Wild-type D39, a mutant with the whole islet integrated into the chromosome (CH226) and a strain isogenic to CH226 but with the *rlrA* gene inactivated (CH232) were used in competition assays. Three experiments were performed in each of two murine models (assessing pneumonia and nasopharyngeal carriage). D39 was competed against CH226 or CH232 and CH226 was competed against CH232.

CH226 out-competed D39 in both models of infection consistent with acquisition of the locus resulting in increased virulence in a pneumonia model and a more successful carriage phenotype (Figures 6.8 (A) and (B)). This gain in function appeared to be due to the presence of the islet as inactivation of the *rlrA* gene encoding the transcriptional regulator reversed the phenotype (Figures 6.8 (A) and (B)).

6.2.5 *In vivo* transfer of the islet

The nasopharynx contains many different bacterial species including pathogenic and non-pathogenic organisms (Kononen *et al.*, 2002). There is likely to be inter-species and intra-species competition for colonization and growth. The nasopharynx potentially provides an environment conducive to inter- and intra-species transfer of genetic material. The *rlrA* islet may have been acquired from another Gram-positive species or another strain of *S. pneumoniae* resulting in increased fitness with respect to colonizing ability. The presence of the *rlrA* islet in only a small subset of *S. pneumoniae* suggests

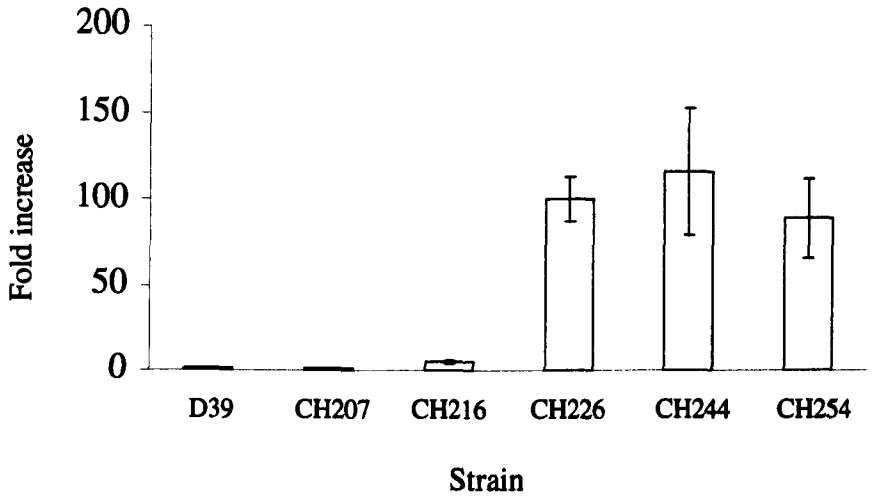


Figure 6.7 Effect of MgrA on adherence of serotype 2 *S. pneumoniae* strains to A549 lung epithelial cells

A549 monolayers were inoculated with D39 (wild-type), CH207 (*ΔmgrA*), CH216 (*mgrA*), a serotype 2 strain possessing the locus CH226 (*IS1167::* magellan5) and *mgrA* mutant strains of CH226, CH244 (CH226 *ΔmgrA*) or CH254 (CH226 *mgrA*$^{++}$) *S. pneumoniae* strains. Mid-exponential growth phase bacterial cells (1x10^6 CFU) were added to A549 cells at a multiplicity of infection of 10:1 and incubated for 40 minutes. The percentages of bacteria that adhered to the monolayers after three washes with PBS were determined. Triton/EDTA was added to each well to detach the cells and liberate bacteria, which were then enumerated by plating lysates on blood agar plates. The adherence of each strain is expressed as a fold increase relative to binding of wild-type D39 cells. Representative results from one of three independent experiments are shown. The assay was performed in quadruplicate in each experiment and the error bars represent the standard deviations.

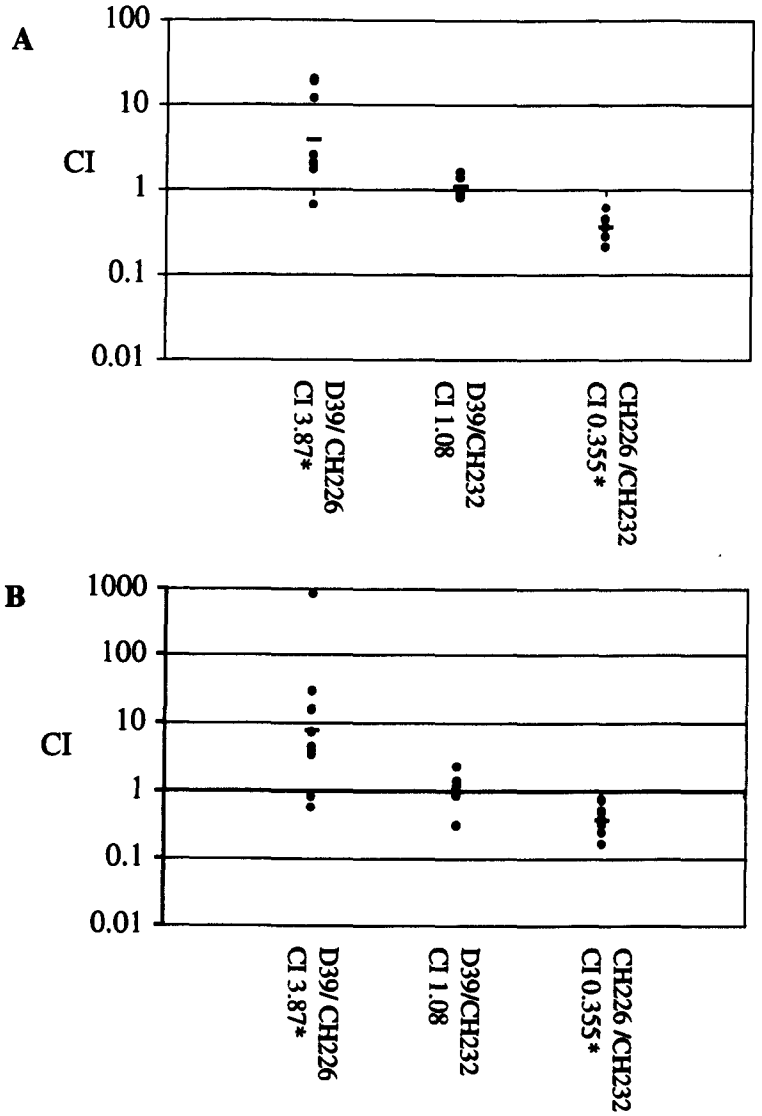


Figure 6.8 Competition assays for D39 strains - (A) pneumonia and (B) Nasopharyngeal carriage

Competition assays were performed using a murine model of pneumonia. Wild-type D39, a mutant with the whole islet integrated into the chromosome (CH226) and a strain isogenic to CH226 but with the *rlrA* gene inactivated (CH232) were used. The *in vivo* competition index (CI) was calculated; each circle represents the CI for a single mouse in each set of competitions. A CI of less than 1 indicates a virulence defect, a CI of greater than 1 indicates that the mutant out-competes the wild-type strain and a CI of 1 indicates that both strains are equal. The geometric mean of the CIs for all mice in a set of competitions is shown as a solid line. The CI marked with an asterix were statistically different to 1.0 (*$p<0.05$).

that it may have been acquired late in the evolutionary history of the organism or that it does not confer a selective advantage to all strains during interactions with the human host. To determine if inter-species transfer of the *rlrA* islet can occur *in vivo*, mixed infections in a murine model of carriage were performed. Mice were inoculated with a combination of two *S. pneumoniae* strains; one serotype 4 strain possessing an antibiotic-resistant marked islet (Spc^R) and a second differently antibiotic resistant-marked serotype 2 strain (Sm^R). This second strain does not possess the islet. Five mice were inoculated with ~ 5 x 10^7 CFU on day 0 with a 1:1 ratio of the serotype 4 and serotype 2 strains. On day 7 they were euthanased and bacteria harvested from nasopharyngeal washes. Serial dilutions of these washes were plated onto blood agar plates supplemented with streptomycin (Sm) 100 μg ml^{-1} or spectinomycin (Spc) 200 μg ml^{-1} or both antibiotics. Acquisition of the islet by the serotype 2 strain or acquisition of Sm^R by the serotype 4 strain would be detected by a Spc Sm resistant phenotype, islet presence was confirmed by PCR and serotype of the Spc Sm resistant strain was determined by Quellung reaction. A competitive index for colonization was calculated as well as the proportion of any Sm Spc resistant bacteria. The serotype 2 strain was completely out-competed by the serotype 4 strain (CI < 0.001). No Sm^R Spc^R colonies were obtained from three of the mice. 2 and 3 doubly resistant colonies were obtained from each of the other two mice, equating to ~ 1 per every 10^4 – 10^5 bacteria in the total output. However each isolate was shown to be serotype 4 by Quellung reaction. This phenotype could result from spontaneous point mutation in the native *rpsL* gene resulting in streptomycin resistance, or from *in vivo* transfer of the mutated *rpsL* gene from serotype 2 to serotype 4 strains.

Given that the serotype 4 strain was more successful in colonizing than the serotype 2 strain a second experiment with altered input ratio in favour of the serotype 2 strain was performed. Five mice were inoculated with ~ 5 x 10^7 CFU with an input ratio of 10:1 (serotype 2:serotype 4). No doubly resistant colonies were isolated on the output plates. Colonies from both serotypes were isolated but the serotype 2 was again out-competed 150 fold by the serotype 4 strain (CI = 0.0064). There was no evidence of inter-strain transfer of the islet in the experiments described above.

6.2.6 Distribution of the islet amongst clinical isolates and correlation with invasion and carriage

An initial screen has shown that the *rlrA* islet is present in only a proportion of clinical isolates of *S. pneumoniae* (Figure 5.6). Given that three of the islet genes are known to encode virulence factors and experiments described above indicate that transfer of the islet into a strain that does not normally harbour it leads to a gain in function as assessed by binding to A549 cells and murine models of pneumonia and carriage I wanted to examine the distribution of the islet in clinical isolates and to determine if there was any correlation between presence of the islet and virulence.

A collection of 501 *S. pneumoniae* isolates from children <5 years old in Oxford has been built up over 8 years (1995-2003) (Brueggemann *et al.*, 2003). One hundred and fifty isolates were obtained from children with invasive disease and 351 isolates are from three longitudinal studies of pneumococcal carriage. This collection has been previously characterised for sequence type by multilocus sequence typing (MLST) (Brueggemann *et al.*, 2003; Enright and Spratt, 1998; Meats *et al.*, 2003). Genomic DNA was obtained from a selection of isolates from this collection. These were

screened for the presence of islet by dot blot hybridization using a control probe to *rpoB* and a test probe to *rlrA*. The *rpoB* probe was used as a control to demonstrate the presence of non-significantly divergent DNA. The same strains were also tested by PCR for the presence of both these genes and two other islet genes, *srtD* and *rrgA*. The samples were allocated random numbers in Oxford to avoid any knowledge of serotype affecting interpretation of the dot blot and PCR results. These were un-blinded after completion of all the experiments. There was an excellent correlation between the dot blot and PCR results. There were no discrepant PCR results consistent with the presence of the primer sequences in each gene among different isolates.

The data from the invasive strains, representative of the serotypes covered by the pneumococcal conjugate vaccine (Figure 5.6), indicated that the locus is only present in a proportion of serotypes. As only one strain in each serotype was examined it was not possible to say whether islet presence was serotype-specific or randomly distributed across strains of all serotypes. To test the hypothesis that islet presence was indeed serotype-specific clinical isolates of serotypes 4, 9V and 9N were examined more closely. The Oxford collection contained six serotype 4 strains (5 isolated from children with invasive disease and 1 from nasopharyngeal carriage) and 15 serotype 9V strains (7 invasive and 8 carriage strains), 6 serotype 9N. These strains were selected for study. 9N is genetically closely related to 9V, strains (1 invasive and 5 carriage) and were taken for comparison. Examination of these isolates would also help determine if islet presence was associated with invasion or not. All serotype 4 strains were positive for islet presence, no serotype 9N haboured the islet and 13 out of 15 9V isolates were positive for the islet. The two 9V strains that did not habour the islet had the same MLST sequence type (ST), which was different from the rest of the 9V isolates.

Table 6.1 Islet distribution amongst *S. pneumoniae* isolates: serotypes 4, 9V and 9N

Serotype	ST[a]	Total[b]	Invasive[b]	Carriage[b]	Islet presence
4	205	4	3	1	Y
	206	1	1		Y
	246	1	1		Y
9V	162	9	6	3	Y
	163	2	1	1	Y
	406	2		2	N
	407	2		2	Y
9N	66	6	1	5	N
	405	1		1	N

[a] MLST sequence type

[b] Data are number of isolates

These data are consistent with the hypothesis that islet presence associates with ST. However from previous characterization of isolates in this collection it is known that isolates of closely related STs commonly have the same serotype and hence islet presence would also associate with serotype.

Next a greater number of isolates were screened. Three isolates from each of the 13 most commonly found serotypes (excluding serotypes 4, 9V and 9N) were examined. Isolates were randomly selected for ST and invasive or carriage site collection. Only two further serotypes were found to possess the islet taking the total number of serotypes to five; 4, 6B, 9V, 19F and 38 (Table 6.2).

An empirical odds ratio (OR) was calculated to compare the probability of invasive disease due to strains possessing the islet to those without the islet to determine if there was a correlation between site of isolation of the strain (invasive or carriage) and islet

presence. Results from all isolates or only isolates from serotypes that harboured the islet were analysed. The OR was calculated using a method published elsewhere (Smith *et al.*, 1993). An OR of 1 indicates that the islet was equally likely to be found in an invasive isolate or carriage isolate, an OR >1 indicates an increased probability for the islet to be found in invasive isolates, and an OR <1 indicates a reduced probability for the islet to be found in invasive isolates. Islet presence was not significantly associated with invasive disease (Table 6.3).

Table 6.2 Islet distribution amongst *S. pneumoniae* isolates

Serotype	ST[a]	Total[b]	Invasive[b]	Carriage[b]	Islet presence
1	227	3	2	1	N
3	180	2		2	N
	458	1	1		N
6A	65	1		1	N
	395	1		1	N
	457	1	1		N
6B	138	1	1		Y
	176	1		1	Y
	273	1		1	Y
7F	191	3	2	1	N
8	53	3	1	2	N
14	9	2	1	1	N
	124	1		1	N
15B/C	411	2	1	1	N
	412	1		1	N
18C	113	2		2	N
	120	1	1		N
19A	172	1	1		N
	193	1		1	N
	417	1		1	N
19F	43	1	1		N
	420	1		1	Y
	426	1		1	Y
23F	36	1	1		N
	40	1		1	N
	311	1		1	N
33F	60	3	2	1	N
38	310	2	1	1	Y
	393	1		1	Y

[a] MLST sequence type, [b] Data are number of isolates. Serotypes that possess the islet are highlighted in grey

Table 6.3 Association of islet presence and invasive disease

	Invasive[a] (%)	Carriage[a] (%)	OR[b]	CI[c]	P value
All strains (n=70)	14 (47)	13 (33)	1.8	(0.6-5.4)	0.3[d]
Strains with islet (n=30)	14 (93)	13 (87)	2.2	(0.1-68)	0.5[e]

[a] Data are numbers of isolates that possess islet. The percentage of the total carriage or invasive strains is reported in parentheses.[b] Odds ratio (OR). [c] Cornfield 95% confidence interval. [d] Yates corrected test of significance used. [e] Fisher exact test of significance used

6.3 SUMMARY

In this chapter several approaches were taken to investigate the physiological significance of *rlrA* islet presence. Presence of the islet in TIGR4 was associated with an ability to bind to A549 cells *in vitro*. This binding was dependent on expression of islet genes and more specifically on the presence of *rrgA*. The islet could be transferred to D39 *in vitro* and resulted in integration into the genome at the equivalent site to that which it is found in TIGR4. Integration into the genome was sufficient to result in a change in phenotype of the *S. pneumoniae* strain to one of increased binding to A549 cells and also increased carriage ability and virulence in murine models of carriage and pneumonia. The hypothesis that integration into the genome was also sufficient to allow regulation of islet genes by endogenous chromosomal based transcriptional regulators in addition to the islet encoded regulator was tested by inactivation of *mgrA* in the islet transformed D39 background but results were inconclusive. *In vivo* transfer of the islet in a murine model of carriage could not be demonstrated. Finally the distribution of the islet amongst clinical isolates was examined. An empirical odds ratio (OR) was calculated to compare the probability of invasive disease due to strains possessing the

islet to those without the islet to determine if there was a correlation between site of isolation of the strain (invasive or carriage) and islet presence. The results did not reach statistical significance and hence I was unable to obtain convincing evidence for any association.

CHAPTER 7 DISCUSSION

Analysis of virulence factors used by *S. pneumoniae* is essential for our understanding of the molecular mechanisms underlying the pathogenesis of disease. Identification and functional analysis of virulence genes and their protein products not only furthers our general understanding of mechanisms involved but can also provide us with specific knowledge that can be used in the development of antimicrobials or vaccines. An important goal would be to understand the timing and signals for expression and activation of every virulence factor and how virulence gene expression is co-ordinated in relation to other cellular processes. In this way it would become apparent which are the crucial steps to invasive disease and hopefully pin-point areas that are amenable to manipulation with a view to preventing and treating disease. With advances in molecular biological tools and techniques it is conceivable that this type of goal is attainable.

The first part of this study describes attempts to adapt a recombination- based *in vivo* expression technology (RIVET) for use in *S. pneumoniae* with the view of examining the timing of virulence gene expression. The latter parts of the study focus on the investigation of the roles of specific transcriptional regulators in virulence gene expression. The work provides further evidence that virulence determinants are not evenly distributed amongst strains of *S. pneumoniae*. It appears that different strains vary both in their quota of virulence factors and their regulation.

7.1 *S. PNEUMONIAE* AND RIVET

At the time of initiation of this project our knowledge of expression of virulence factors and their regulation in *S. pnemoniae* was minimal. Studies in other pathogens such as *V. cholerae* and *B. pertussis* had given us some evidence that virulence factors show distinct patterns of expression in terms of timing and site of expression *in vivo* (Lee *et al.*, 1999) and that the exact timing is important if not crucial to appropriate disease progression (Akerley *et al.*, 1995). I hypothesized that this is similar in *S. pneumoniae* and with the goal of examining the expression of a known virulence factor, pneumolysin, I attempted to adapt the RIVET system for use in *S. pneumoniae* using a *ply::tnpR* transcriptional fusion. This would have enabled me to determine the timing of *ply* gene induction *in vitro* initially and then during the course of a murine infection.

Preliminary experiments showed that *ply* transcription was regulated over the course of *in vitro* growth, as determined by measurement of transcript levels (Figure 3.2). *ply* transcript was detectable even at early stages of growth and there was an increase in levels during exponential phase of growth. This was consistent with previous findings that cytoplasmic and extracellular Ply activity during growth *in vitro* is not normally detectable until late log phase (Benton *et al.*, 1997). However this result could not be repeated by screening for resolution over the course of *in vitro* growth with RIVET constructs using a *ply::tnpR* ^{mut135} reporter in an *S. pneumoniae* strain that possessed the resolvase substrate cassette elsewhere on the chromosome. Resolution was not detectable *in vivo* during the course of a murine pneumonia infection either.

Two possible reasons for failure of the initial RIVET constructs in *S. pneumoniae*, could have been due to a lack of a resolvable cassette or insufficient resolvase

production. The cassette was shown to be functional in an *E. coli* background (Figure 3.4) and there was read-through transcription from the *ply* gene to *tnpR*^{mut135} as assessed by presence of a *tnpR* transcript in RNA extracts from the *S. pneumoniae* fusion strain (Figure 3.5). Attention was turned to assessment of production of the resolvase protein. One of the main drawbacks of IVET and RIVET as they were first described is that to detect induction of transcription *in vivo* any transcriptional fusion needs to be silent *in vitro*. Thus genes that are transcribed *in vitro* will be excluded from use even if they are further induced *in vivo* and play a pivotal role in infection. The beauty of the resolvase based version of IVET (or RIVET) is that it has shown that it can be adapted to overcome the problem of baseline transcription *in vitro* by alteration the translational efficiency of the *tnpR* mRNA (Lee *et al.*, 1999; Lee *et al.*, 2001). The RIVET system relies on production of sufficient resolvase, encoded by *tnpR*, to act on its substrate, an antibiotic resistance gene-containing substrate cassette, resulting in excision of a cassette from the genome (Figure 3.1). If a *tnpR* allele with a less efficient RBS is used in the fusion, less resolvase is produced for any level of transcription. There will be a threshold at which increased transcription will result in adequate enzyme production to give resolution, but below which there is insufficient enzyme to allow resolution. As there was detectable *ply* transcription during *in vitro* growth the initial *ply S. pneumoniae* reporter strain was constructed with a *tnpR* allele (*tnpR*^{mut135}) with reduced translational efficiency. It was possible that this reduced translation to such an extent that insufficient enzyme was produced to support resolution in the conditions tested. Supporting this hypothesis was the evidence that there was no detectable resolvase protein present in cell extracts from *in vitro* grown *S. pneumoniae* fusion strains (Figure 3.5). This was not alleviated by replacement of the mutant *tnpR*^{mut135} with a *tnpR* allele

that possessed an improved RBS (Figure 3.6) indicating that the lack of translation was not secondary to an inadequate ribosomal binding to initiate transcription.

RIVET was initially demonstrated in a Gram negative bacterial background (Camilli *et al.*, 1994; Camilli and Mekalanos, 1995). The resolvase gene was isolated from the Gram negative bacterial transposable element Tn$\gamma\delta$, which is one of a family of transposons that are related to Tn*3* (Heffron, 1983). Although the majority of studies using the system had been in Gram-negative pathogens, it had been used successfully to identify *in vivo* induced genes in one Gram-positive organism, *Staphylococcus aureus* (Lowe *et al.*, 1998). In this study the RBS site of the *tnpR* had to be replaced with a Gram-positive homologue to improve the translation of the resolvase in *S. aureus*. It was surprising therefore that this technique had not worked in *S. pneumoniae*. Codon usage has been shown to be important in determining translation rates in bacteria (Sorensen *et al.*, 1989). Analysis of the codons in *tnpR* compared to usage within the pneumococcal genome identified 20 out of the total 184 codons of *tnpR* that were atypical and rarely used in *S. pneumoniae* (Figure 3.7). This number could be sufficient to abolish *tnpR* mRNA translation. In *E. coli* abolishment of rare codons and replacement with more common ones can alter mRNA translation to such a degree as to eliminate some post-transcriptional regulatory effects (Kuhar *et al.*, 2001). Alteration of the nucleotide sequence of the *tnpR* gene to abolish rare codons throughout the gene and replace them with more frequently used ones resulted in detectable resolvase protein in whole protein extracts from *in* vitro grown strains of S. *pneumoniae* (Figure 3.10). This is consistent with the hypothesis that the codon usage in the native *tnpR* gene was not optimal for translation in *S. pneumoniae*. Yet despite the presence of resolvase in quantities comparable with those seen in AC311 (Figure 3.10), a *V. cholerae* strain

harbouring *vieS'::tnpR* fusion and a *res-tet-res* cassette, that showed 100% resolution, resolution was only detectable in 30% of the *S. pneumoniae tnpR* fusion strains (Figure 3.11). Levels of resolution *in vivo* were even less (Figure 3.11).

Two forms of the protein were repeatably recognised by the anti-resolvase antibody in protein extracts from the *tnpR* fusion strains (Figure3.9) that were not clearly present in purified resolvase preparations or in AC311 (Figures 3.5, 3.6 and 3.10). The reason for this is unclear. Resolvase is a 21 kDa protein that is composed of two structurally and functionally distinct domains. The C terminal domain has DNA binding activity and binds to sequences within the *res* sites. The N terminal domain mediates multimerisation of the enzyme. The active form of the enzyme usually consists of a dimer (Hatfull and Grindley, 1988). Both the monomer and dimer were apparent on SDS–PAGE gels of the purified protein (data not shown). The second protein recognized by anti-resolvase antibodies in the *S. pneumoniae tnpR* fusion strains was between 22 and 25 KDa in mass and as such did not represent a dimer. It is not known if this higher molecular weight protein represents an abnormal non-functional resolvase secondary to translation initiation at an alternative start codon, post-translational modification or purely an artifact of the gel production. If it is a non-functional resolvase-like protein that was able to bind *res* sites and/or dimerise with functional resolvase, thus blocking the action of the usual form of the enzyme, it is conceivable that this dominant-negative activity could explain the lack of high levels of resolution seen in the *S. pneumoniae* fusion strains. Some credence is given to this hypothesis by the observation that increasing the amount of resolvase present, as determined by Western blots (Figure 3.10) resulted in a trend towards reduction in amount of resolution (Figures 3.11A). Though not statistically significant, there was also a trend

towards decreasing levels of resolution with increasing concentrations of maltose in the same strain (Figure 3.11) but it is not clear if this is secondary to increasing levels of TnpR expression or an independent affect of increasing maltose concentrations. Increasing TnpR could cause a reduction in resolution if the ratio of dominant negative to wild-type TnpR were to increase. It would be interesting to perform Western blots for resolvase enzyme on protein extracts from the *MalM::tnpR*^{M} fusion strain (CH102) grown in increasing concentrations of maltose and quantitate the dose-response. This would confirm whether a reduction in resolution is associated with more or less resolvase production. If the former were true then this would be further evidence for a dominant negative effect from a second form of the enzyme.

The work presented in chapter one raises as many questions as it answers with respect to adaptation of RIVET for use in *S. pneumoniae* and reasons for failure. More experiments could be designed to investigate the problem further but from the perspective of the project and aim of investigating expression of virulence genes in this organism it was decided that this would not be beneficial to proceed with these experiments as there was no guarantee that a useful system could be developed, and hence work on RIVET was ceased.

7.2 CHARACTERISATION OF THE VIRULENCE DEFECT SEEN WITH INACTIVATION OF TRANSCRIPTIONAL REGULATORS IDENTIFIED BY STM

The second part of this study focuses on transcriptional regulators identified by an STM screen and investigation of their role in virulence in murine models of *S. pneumoniae* disease. There is good evidence that regulation of gene transcription is important in virulence in *S. pneumoniae*. This is not surprising, as one would predict that the pathogen would have mechanisms to control virulence factor expression at the levels of gene transcription and translation, as well as regulating post-translation modifications to alter protein activation status.

Evidence for the importance of transcriptional regulators in virulence in *S. pneumoniae*, that was available at the conception of this thesis project, came from STM screens and work following on from genomic searches to identify TCSTS in this pathogen. Three large scale STM screens in *S. pneumoniae* had been published (Hava and Camilli, 2002; Lau *et al.*, 2001; Polissi *et al.*, 1998). Two report the identification of factors involved in regulation of gene transcription that are required for virulence in murine models (Hava and Camilli, 2001; Lau *et al.*, 2001). Two separate genomic searches had identified the genes in the *S. pneumoniae* encoding 13 TCSTS genome, many of which were also highlighted in the STM screens, and assessed their role in virulence. None were required for virulence in serotype 3 or 22 *S. pneumoniae* (Lange *et al.*, 1999) in a septicaemia model, but eight have been shown to be required for full virulence in serotype 3 *S. pneumoniae* in a murine model of pneumonia (Throup *et al.*, 2000) and of these one is also required for efficient nasopharyngeal carriage in newborn rats in

serotype 3 *S. pneumoniae* (Sebert *et al.*, 2002). I hypothesized that genes identified in the STM screens that were predicted to encode proteins with a regulatory function, either isolated transcription regulators or TCSTS, controlled the transcription of known virulence genes and yet unidentified virulence genes.

I confirmed that 5 genes encoding probable transcriptional regulators identified in one of the STM screens (Hava and Camilli, 2002) are indeed required for virulence as assessed by competition assays against wild-type bacteria (Table 4.1). Three genes, *sp1800* or *MgrA*, *sp0461* or *rlrA*, and *sp0247*, encode transcriptional regulators. Two, *sp0661* and *sp0156*, encode components of TCSTS. Mutant strains with deletions or transposon insertions in each gene under study were tested in three different murine models; pneumonia, bacteraemia and nasopharyngeal carriage. The mutants show varying attenuation in the three models. CH107 (*sp0247::magellan2*) was attenuated to some degree in all 3 models, CH119 (Δ*sp0661*) is attenuated in lung and bacteraemia only, AC1213 (*rlrA::magellan2*) and AC1500 (Δ*mgrA*) in lung and carriage only and CH108 (*sp0156::magellan2*) in lung only. The requirement for virulence, therefore, is dependent on the type of infection and model used. This is consistent with the hypothesis that the regulators control the transcription of genes whose products are required in specific host environments and are important in adaptation to the different environments that the bacteria encounter during host infection or colonization.

These data support the results from the STM screens and the studies on TCSTS, which also found varying attenuation in different disease models. The data for *sp0661*, however, is not fully consistent with the results from these published works. This gene encodes the response regulator TCS09 (nomenclature of Lange et al, (Lange *et al.*,

1999)). In the thesis work presented here, the serotype 4 *S. pneumoniae sp0661* deletion-mutant was attenuated in both a pneumonia and a septicaemia model (Table 4.1). The mutant was not impaired in *in vitro* growth (data not shown). Lange and colleagues created knockout mutations in the genes encoding the same response regulator of TCS09 of serotype 3 and serotype 22 *S. pneumoniae* strains and found no difference in virulence between the mutant and wild-type strains in a similar murine septicaemia model. The infecting dose was similar in both studies but the strain of mouse, C57BL6/J (Lange *et al.*, 1999) or Swiss Webster (this work), and the *S. pneumoniae* serotypes were different. The genetic background of the bacterial strain or the mice, or a combination of both may have resulted in the discrepancies observed (Gingles *et al.*, 2001). The former certainly appears to be plausible. The contribution to the virulence of *S. pneumoniae* of the gene encoding the same response regulator (*rr09*) was shown to be bacterial strain-dependent in a further study (Blue and Mitchell, 2003). Here *rr09* was found to be essential for virulence in D39 (serotype 2 strain) but not in a serotype 3 strain, when inoculated into MF1 mice by the intra-peritoneal route. This strain difference suggests that the TCSTS, Rr09/Hk09, has different gene targets in genetic backgrounds.

However there may be another explanation. The original serotype 4 *sp0661* transposon-insertion mutant (STM90) was attenuated in a pneumonia model but did not show attenuation in a septicaemia model (Table 4.1 and (Hava and Camilli, 2002)). The transposon was not within the open reading frame of *sp0661* but 21bp upstream of the predicted ATG start codon. It is possible that this transposon insertion disrupts *sp0661* expression resulting in either expression at inappropriate times or in inappropriate quantities. Alternatively the transposon insertion may exert polar effects on other genes

in the operon, including *sp0662*, encoding the histidine kinase component of the TCSTS. The deletion mutant used in this thesis work had an unmarked, in-frame deletion resulting in fusion of the ATG start codon to the last 206 bp of the *sp0661* open reading frame. This 206 bp contained the RBS and first 6 bp of the open reading frame for *sp0662*, Thus, the unmarked deletion mutant should result in preservation of *sp0662* transcription and translation. Although the genes encoding this response regulator (RR) and the histidine kinase (HK) lie adjacent to each other on the chromosome and are predicted to be part of the same TCSTS it is possible that their transcription is differentially regulated, with a promoter lying within the coding region of *sp0661*, or that the HK, once activated, could activate another RR aside from SP0661. Similarly SP0661 could be activated by a second HK. There are several examples of such cross talk between TCSTS (Li *et al.*, 1995; Msadek, 1999). Disruption of both components may therefore have an overall different effect than disruption of one alone and could explain the variation in virulence phenotype between different mutants.

The two strains with transposon insertions in or around *mgrA* from the STM screen (STM206 and AC1272) and the *mgrA* deletion strain from this work (AC1500) also had slightly different phenotypes in competition assays (Figure 4.2 and Table 4.1). A transposon insertion 300 bp upstream of the start site (STM206) had the greatest effect on virulence. RPA (data not shown) probing for *mgrA* transcript in the three mutants showed that there was no detectable *mgrA* transcript in either the deletion mutant or mutant with the transposon inserted into the coding sequence, but a transcript of a larger size than that seen in the wild-type strain was detectable in the mutant with the transposon insertion upstream of the *mgrA* start site. It is possible that in this latter strain the transposon insertion leads to disregulation of *mgrA* due to read-through

transcription from the transposon or effects on protein binding to the *mgrA* promoter region. This hypothesized inappropriate gene expression may be more detrimental than an actual gene knockout. Alternatively, I have not excluded the possibility that the upstream transposon insertion also has effects on the putative open reading frame in the opposite direction, *sp1801-1805*, and that these gene products are required for virulence (Figure 4.2).

7.3 RLRA IS A TRANSCRIPTIONAL ACTIVATOR PRESENT ON A PATHOGENICITY ISLAND

Having assessed the requirement for different transcriptional regulators in murine infection, experiments were performed to identify which virulence genes they were controlling. Transcriptional regulators often lie adjacent to their target genes on the chromosome but they may also regulate distant genes. The mutants with transposon insertions in genes *sp0461* and *sp1800*, renamed *rlrA* (encoding RofA-like regulator A) and *mgrA* (encoding Mga-like repressor A), are attenuated in models of pneumonia and nasopharyngeal carriage. Neither appear to be essential for virulence in a septicaemia model, suggesting that the products of these two genes are required for regulation of factors that interact with the mucosal surfaces of the respiratory tract.

Both of these genes are predicted to encode proteins with similarity to transcriptional regulators in GAS (*S. pyogenes*). *Sp0461* encodes a protein similar to RofA and Nra (Fogg *et al.*, 1994; Podbielski *et al.*, 1999) and *sp1800* encodes a protein that is similar to the multiple gene regulator, Mga (McIver and Scott, 1997; McIver and Myles, 2002). In GAS Mga and RofA/Nra regulate transcription of virulence genes encoding primarily

surface-associated products, many of which lie adjacent to the regulator on the chromosome.

Analysis of the *S. pneumoniae rlrA* mutant using RPAs revealed that it is a transcriptional activator of itself and neighbouring genes on the chromosome (Figure 4.5 and (Hava *et al.*, 2003a)). The products of the target genes are predicted to be cell surface-associated and as such capable of affecting bacterial-host interactions at mucosal surfaces. *RlrA* is divergently transcribed from three genes encoding cell wall anchored proteins and three encoding sortases. The gene organization is shown in Figure 4.1. RPAs were used to measured steady state levels of each of these six adjacent genes in wild-type and *rlrA* mutant strain. The message was found to be decreased in the mutant strain as compared to the wild-type (Figure 4.5). The level of decrease varied for each gene in a pattern consistent with the hypothesis that RlrA binds to and regulates three promoters, one controlling *rrgA*, a second *rrgBC* and a third *srtBCD*. This has subsequently been confirmed by primer extension analysis and DNAse I foot-printing (Hava *et al.*, 2003a).

The organization and functional analysis of this 12-kb region of DNA containing *rlrA* and its targets indicates that it represents a PAI in *S. pneumoniae*. A PAI is minimally defined as a multi-gene region of the genome, which is absent or sporadically distributed in non-pathogenic close relatives and that carries one or more virulence factors. The overall nucleotide composition of the island often differs from the core genome in G+C content and codon usage, reflecting the acquisition of PAIs by inter-species horizontal gene transfer. PAIs are often flanked by sequences defining the edges, such as direct repeats that could arise from integration of a transposon or

bacteriophage (Hacker *et al.*, 1997; Hacker and Kaper, 2000). The *rlrA* region fulfils some of these criteria. *RlrA* is divergently transcribed from the six genes that it regulates (*sp0462 – sp0468*). This whole region is flanked by two IS1167 insertion sequences, which defines the islands edges. Three of the seven genes, *sp0461*, *sp0462* and *sp0468*, are required for virulence in serotype 4 *S. pneumoniae* (Hava and Camilli, 2002; Hava *et al.*, 2003a) and the island is absent in many *S. pneumoniae* strains [(Tettelin *et al.*, 2001) and Figure 5.7]. The G+C content of the island is not significantly different from the *S. pneumoniae* genome as a whole but a G+C difference between island and genome may not be observed if the DNAs of the donor of the PAI and the recipient have a similar G+C content. It is possible that the island was acquired by *S. pneumoniae* from another streptococcal species such as *S. agalactiae*. Sequence similarity searches using the BLAST program (http://www.ncbi.nlm.nih.gov/blast/) identified a region in *Streptococcus agalactiae* that had a similar organization, namely a gene encoding transcriptional regulator divergently transcribed from three encoding cell wall anchored proteins and three encoding sortases. Many of these genes (*SAG0644 - SAG0649*) share small regions of sequence similarity to *sp0461 - sp0468* of *S. pneumoniae*.

One other pathogenicity island, PPI1, has been formally described in *S. pneumoniae* (Brown *et al.*, 2004; Brown *et al.*, 2001). This 27-kb region contains *piaABCD*, encoding an iron transporter, and 28 other genes, some of which have been shown to encode virulence factors (Brown *et al.*, 2004). The *rlrA* island will therefore be referred to as PPI2. There are undoubtably more PAIs to be identified and characterised in this pathogen. As different *S. pneumoniae* serotypes differ by as much as 10% in gene composition, PAIs may in fact compose a major fraction of the genome.

7.4 MGRA IS A TRANSCRIPTIONAL REPRESSOR OF GENES OF PPI2.

Mga, of GAS, activates the transcription of several genes encoding cell surface-associated virulence factors involved in *S. pyogenes* pathogenesis, including the M family of proteins, C5a peptidase, serum opacity factor and secreted inhibitor of complement (Caparon and Scott, 1987; Chen *et al.*, 1993; Kihlberg *et al.*, 1995; McLandsborough and Cleary, 1995). The genes encoding M protein and C5a peptidase (McIver and Myles, 2002; Pobielski *et al.*, 1996; Rasmussen *et al.*, 2000; Terao *et al.*, 2001) lie adjacent to *mgA* on the GAS chromosome. RPAs were used to determine if *S. pneumoniae* MgrA affected the transcription of neighbouring chromosomal genes and microarray experiments were performed as part of a wider search for genes under MgrA control.

Analysis of the *S. pneumoniae mgrA* deletion and over-expressing strains using RPAs revealed that MgrA did not effect transcription of neighbouring genes in the chromosome *in vitro* (Figure 4.6 and (Hemsley *et al.*, 2003)). Microarray experiments comparing mRNA levels from *in vitro* grown wild-type *S. pneumoniae* to those from same *mgrA* mutants, however, identified that the level of transcript for each gene on PPI2 showed a consistent and statistically significant difference of 2-5 fold between strains at all growth time-points assayed. Transcripts were increased in the *mgrA* deletion strain compared to the *mgrA* over-expressing strain (Figure 5.2 and Table 5.1), consistent with the notion that MgrA acts as a repressor of transcription of genes within PPI2. This is perhaps not altogether surprising as it is known that the similar regulators in GAS show cross-regulation, with Nra acting as a transcriptional repressor of *mga* whilst in turn Mga activates *nra* transcription (Pobielski *et al.*, 1999). There was no

reciprocal regulation of *mgrA* by RlrA detected in *S. pneumoniae* (Figure 5.4). I propose a model where MgrA binds to the promoter region of *rlrA* to repress *rlrA* transcription. This has the effect of also reducing transcription of other PPI2 genes, as they require RlrA for transcription (Figure 7.1).

It cannot be excluded that MgrA acts at each of the 4 promoters of PPI2 but the fact that deletion of *mgrA* in a *srtD::lacZ rlrA::magellan2* background does not result in an additional increase in β-galactosidase activity over that seen in the *srtD::lacZ rlrA::magellan2* strain argues against this (Figure 5.11). Further experiments could be performed to determine if MgrA does indeed bind to one or more of the promoter regions in the islet. These include:

a) Gel mobility shift assays using a His_6-tagged version of MgrA and DNA fragments representing each of the 4 promoter regions in PPI2
b) DNase 1 foot-printing to precisely map the sites of His_6-tagged MgrA binding

What is surprising is that no other obvious candidate target genes for MgrA were identified in the array experiments. PPI2 is not present in all *S. pneumoniae* serotypes (Figure 5.7 and Tables 6.1 and 6.2). In contrast *mgrA* has been found in all of 8 serotypes tested (Figure 5.7). One would expect MgrA to have a role in transcriptional regulation of other genes in strains that do not possess the PAI. I would have also predicted that there are additional genes under MgrA control even in strains that do possess PPI2, such as the TIGR 4 strain used in this study.

What are the possible reasons for the failure to identify additional genes? Firstly, some genes may be regulated below the threshold for detection by microarray. Secondly,

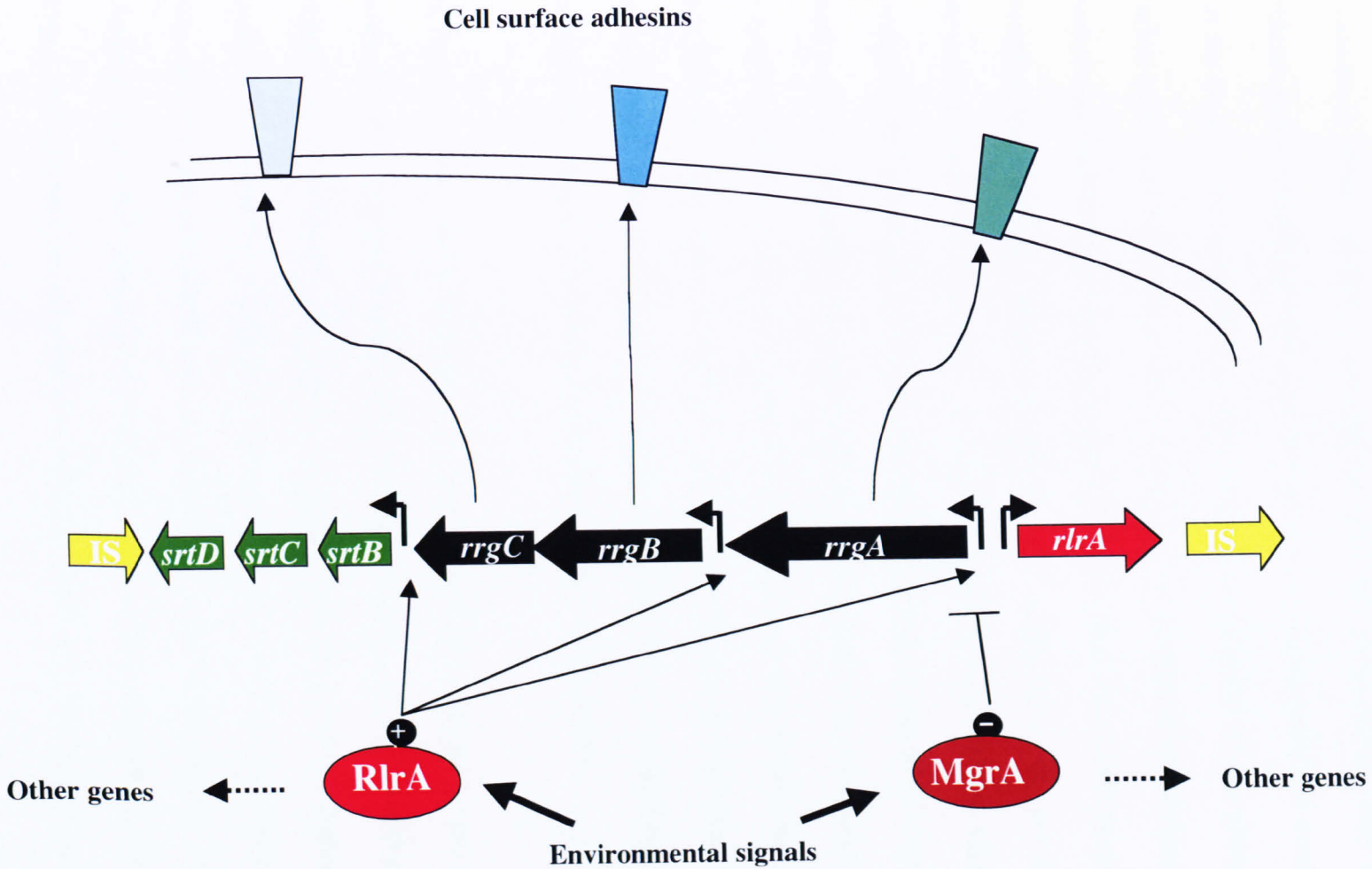


Figure 7.1 Model for transcriptional regulation of PPI2 genes by MgrA and RlrA

MgrA may only regulate genes in a small fraction of cells within a population. Thirdly, it is possible that MgrA does not regulate the expression of other genes in serotype 4 *S. pneumoniae* strains. If this is the case and the PPI2 genes are the only essential virulence genes controlled by MgrA in serotype 4 *S. pneumoniae* then one would expect that the Δ*mgrA rlrA::magellan2* mutant would have a similar *in vivo* phenotype in murine models of infection to the *rlrA::magellan2* mutant strain. Animal studies showed that the COIs were not significantly different from 1 for either pneumonia or nasopharyngeal carriage (Figure 5.12). In other words there was no additional attenuation in mice over and above that seen in the *rlrA* mutant with the addition of a second mutation in *mgrA*, consistent with the idea that both regulators act within the same pathway. Although this suggests that there are no other virulence factors under the control of MgrA in TIGR4 the murine COI experiments do not exclude the possibility that there are other genes, which do not have essential roles in virulence, under the transcriptional control of MgrA. There could be other virulence genes, but ones which act at different stages of infection not simulated in the murine models.

In support of the hypothesis that the only targets for MgrA in TIGR4 are the islet genes is the fact that serotype 2 strains, R6/D39, have a further gene (*Spr1404*) encoding a protein similar to MgrA. This gene is divergently transcribed from a gene (*Spr1403*) predicted to encode for a membrane-associated protein, which has a sortase recognition sequence. Neither *Spr1404* nor *Spr1403* are present in the TIGR4 genome. It is possible therefore that other pneumococcal serotypes have other *mgrA*-like genes and each MgrA-like regulator has only one or a few genes whose transcription it effects, each set being unique to that strain.

Fourthly, MgrA is known to be functional *in vivo* as deletion of the gene results in a change in the virulence properties of the bacteria in murine models. The mRNA for the array experiments was harvested from bacteria grown *in vitro*. Many transcriptional regulators have two forms, one of which has functional activity to regulate gene expression (active form) and a second form, which has reduced or no activity (non-active form). The switch from one form to the other often requires a co-factor or partner protein or a modification such as phosphorylation or dephosphorylation. The *in vitro* conditions used in these studies are unlikely to mimic those seen in a murine infection. The signals required for full regulator activation may not have been present *in vitro* and MgrA may not have been fully active in these conditions. The *in vivo* data from mice (Table 4.1) and the results from the epithelial cell line adherence experiments (Figure 6.4) are consistent with this hypothesis. Strains bearing a deletion of *mgrA* have a decreased ability to cause lung infection and nasopharyngeal colonization despite their increased ability to adhere to lung epithelial cells *in vitro* (Figure 6.4). It is possible that a different set of genes are regulated by MgrA in response to the *in vivo* environment experienced by the *S. pneumoniae* in the mouse than in the *in vitro* conditions in the binding assays. This could lead to expression of different surface factors and different phenotypes with respect to adhesion to host cells.

Even if MgrA does not control genes outside PPI2 in serotype 4 *S. pneumoniae* strains, there appear to be additional genes under its control in other serotypes. There is good evidence that transcriptional regulators of *S. pneumoniae* have varying target genes in different strains. This is the case for the response regulator RR09/SP0661 (Blue and Mitchell, 2003; Lange *et al.*, 1999), MicB (Kadioglu *et al.*, 2003; Throup *et al.*, 2000)

and RegR (Chapuy-Regaud *et al.*, 2003) and data presented here also indicates that this is the case for MgrA. In addition to serotype 4 strains, MgrA is required for virulence of D39 in a murine model of pneumonia but not for nasopharyngeal carriage of the same strain (Figure 5.8). As D39, a serotype 2 *S. pneumoniae* strain, does not contain PPI2 the MgrA regulator must be regulating other virulence factors in this strain. It would be interesting to perform similar microarray experiments to those performed with serotype 4 with comparable D39 strains. Perhaps there are other loci exclusive to D39 that are under the control of MgrA.

7.5 THE FUNCTION AND DISTRIBUTION OF PPI2

PPI2 possesses 7 genes that encode for a transcriptional regulator (RlrA), three surface proteins (RrgABC) and three sortases (SrtBCD) that are required for the anchoring of these surface proteins to the cell wall (Hava *et al.*, 2003a). In serotype 4 and 9V *S. pneumoniae rlrA* and *rrgA* are important for adhesion to lung epithelial cells (Figure 6.2) and I propose that this adhesion is required for nasopharyngeal colonisation and production of pneumonia in the mouse. The *rrgA* gene is predicted to encode for an MSCRAMM suggestive that it has a role in binding to extracellular matrix molecules (ECM). Some other Gram-positive pathogens use binding of ECM molecules to facilitate adherence to and/or invasion of epithelial cells (Massey *et al.*, 2001; Molinari *et al.*, 1997). This is via interactions with integrins on host cells that can bind ECM components. If the ECM component is also bound by the pathogen, then ECM-integrin binding can facilitate bacterial adherence and, in some cases, uptake into the host cell. The actual ligand for RrgA is unknown. RrgA may bind to an ECM component or directly to an epithelial cell receptor. If it does bind an ECM component, such as

fibronectin or collagen, this could act as a bridging molecule between *S. pneumoniae* and epithelial cell integrins.

S. pneumoniae has been notably infrequent in studies describing MSCRAMMs as receptors in Gram-positive pathogens. One such protein has been described in *S. pneumoniae*, PavA, a surface exposed protein that binds fibronectin (Holmes *et al.*, 2001). The *pavA* gene is present in *S. pneumoniae* strains of different serotypes. Determining the ligand for RrgA and defining the ligand-binding domains in the RrgA protein is therefore an interesting area for further work. This would firstly necessitate purification of recombinant RrgA and generation of anti-RrgA antibodies. The raising of antibodies would enable experiments to confirm that RrgA is surface exposed, as predicted. This could be achieved by a) Western immunoblot analysis of proteins present in cytoplasmic, membrane and cell wall fractions prepared from wild-type serotype 4 *S. pneumoniae* compared to D39 and serotype 4 *rrgA* mutants, b) immunoelectron microscopic visualization of RrgA on the bacterial cell surface and/or c) confocal immunofluorescence microscopy. Having confirmed that RrgA is cell surface-exposed, purified protein could also be used in experiments to identify its binding partner. Two broad approaches would need to be taken. Firstly, if RrgA does indeed bind an ECM component then the ligand might be identified successfully in RrgA binding assays to a panel of purified ECM components immobilized onto microtitre plate wells. Secondly, if RrgA binds epithelial cells directly, rather than an ECM component, it may be possible to identify the receptor with the aid of either affinity chromatography, using biotin labeled surface proteins from A549 cells over a RrgA affinity column (Isberg and Leong, 1990; Mengaud *et al.*, 1996), or with the aid of a yeast two-hybrid system.

PPI2 is not present in all clinical isolates but has been acquired by some *S. pneumoniae* strains at some point in their evolutionary history. Are *S. pneumoniae* strains ready primed to accept additional virulence genes, allowing immediate functioning of these genes and integration into already present regulatory networks? Transferring PPI2 into D39, which does not normally possess it, results in increased adherence to A549 cells *in vitro* indicating that the surface adhesins and sortases are successfully expressed, processed and expressed on the bacterial cell surface (Figure 6.6). Acquisition of the element by D39 also results in increased virulence (Figure 6.8). It is apparent therefore that PPI2 contains enough information to function without the apparent requirement for any additional genetic alterations in some *S. pneumoniae* strains that acquire PPI2. Although it could not be demonstrated that transfer of PPI2 occurs *in vivo* in a murine model, *S. pneumoniae* is naturally transformable and recombinational events, such as capsular locus swapping, do occur in the human host *in vivo* (Coffey *et al.*, 1998; Coffey *et al.*, 1999). Such events have played a significant role in the evolution of the *S. pneumoniae* genome (Feil *et al.*, 2000). PPI2 is therefore likely to be transferred between strains *in vivo*.

The fact that PPI2 is found at low frequency amongst clinical isolates (Tables 6.1 and 6.2) indicates that it either does not undergo high frequency transmission between *S. pneumoniae* strains, its acquisition is relatively recent in the evolutionary history of *S. pneumoniae* or that it is not maintained in some strains that acquire it. There could be functional redundancy between PPI2 and other surface adhesins in *S. pneumoniae* strains. The acquisition of PPI2 by some strains may not affect the binding phenotype of

that strain if it already has other adhesins. If there is no selective advantage in having the element there will be no pressure to maintain PPI2 presence in the strain.

Although acquisition of the island by D39 results in increased virulence in murine models of infection, there was no statistically significant correlation between islet presence and invasive disease seen in clinical isolates (Table 6.3).

The genes of PPI2 are regulated by RlrA, encoded by a gene on the island, but also by at least one regulator, MgrA, encoded by a gene present on the chromosome distant from PPI2. The phenomenon of regulators encoded outside a PAI regulating genes encoded on the PAI and vice versa (Kaper and Hacker, 2000). As yet there are no documented examples of chromosomal genes under transcriptional control of RlrA. However a 13-bp consensus sequence binding site for RlrA has been determined (Hava *et al.*, 2003a) and this was used to query the TIGR4 genome sequence for other possible RlrA binding sites. The search identified 29 sequences which were present in putative promoters (Hava *et al.*, 2003a). These include 2 known virulence genes (A. Camilli and D. Hava, personal communications) raising the possibility that RlrA does regulate genes outside PPI2. The hypothesis that RlrA does indeed regulate the transcription these 29 genes could be tested by RPAs. It would also be interesting to perform microarray studies with *rlrA* mutant strains to identify any other genes under the transcriptional control of RlrA.

7.6 ENVIRONMENTAL SIGNALS REQUIRED FOR TRANSCRIPTIONAL REGULATOR ACTIVATION

The gaseous composition experienced by *S. pneumoniae* varies with anatomical site within the human host (Table 7.1). The composition of air at the nasopharyngeal mucosa is higher in oxygen and lower in carbon dioxide than alveolar air or in the middle ear space. The availability of oxygen is also decreased in the common manifestations of disease caused by *S. pneumoniae*, such as pneumonia, empyema and otitis media as the exudative inflammatory response occludes air spaces limiting ventilation and gas exchange (Harell *et al.*, 1996; Hergilis and Magnuson, 1997; Sade *et al.*, 1995).

Table 7.1 Gaseous composition at mucosal surfaces within the human host

	Gaseous composition (%)		
	N_2	O_2	CO_2
Atmospheric/inhaled air	74.09	20.84	0.04
Middle ear[a]	83.66	6.86	9.47
Alveolar Air[a]	74.9	13.6	5.3
Expired air	74.5	15.7	3.6

[a] The figures given are in a non-disease state. The actual composition in compartments, such as the alveolar compartment or middle ear, will also vary with the normal ventilatory cycle and in the presence of pathology (Harell *et al.*, 1996; Hergilis and Magnuson, 1997; Sade *et al.*, 1995).

S. pneumoniae might be able to sense these changes and as a result induce changes in the expression of surface structures affecting interactions with the host epithelial cells and bacterial adhesion. There is some evidence that the ambient oxygen concentration is an important factor in the ability of *S. pneumoniae* to regulate the characteristics of its cell surface (Figure 7.2). The *S. pneumoniae* colony morphology is dramatically different with growth on solid media anerobically as compared to growth non-anaerobically with a shift to a larger and more mucoid colony under anaerobic growth (Weiser *et al.*, 2001). This reflects, amongst other things, a difference in capsular polysaccharide production. Oxygen levels affect the opaque and transparent phase variants differently, with greatest effect is on the opaque (O) variant. There is an increased production of capsular polysacharride (CPS) in O variants in conditions of reduced oxygen, whereas synthesis of CPS in transparent variants remains comparatively low under atmospheric or reduced oxygen conditions (Weiser *et al.*, 2001). The expression of capsule is necessary for *S. pneumoniae* to evade opsonophagocytosic clearance during infection but may be detrimental to adhesion and carriage.

Little is known about the role of carbon dioxide in regulating *S. pneumoniae* cellular processes and surface and other virulence factors although the pathogen requires carbon dioxide for growth (Flanagan and Paull, 1998). In the study by Weiser and colleagues cited above alteration of carbon dioxide levels had no further effect on CPS production over and above that seen with reduction of oxygen levels, consistent with oxygen rather than carbon dioxide being the major determinant in the levels of CPS (Weiser *et al.*, 2001).

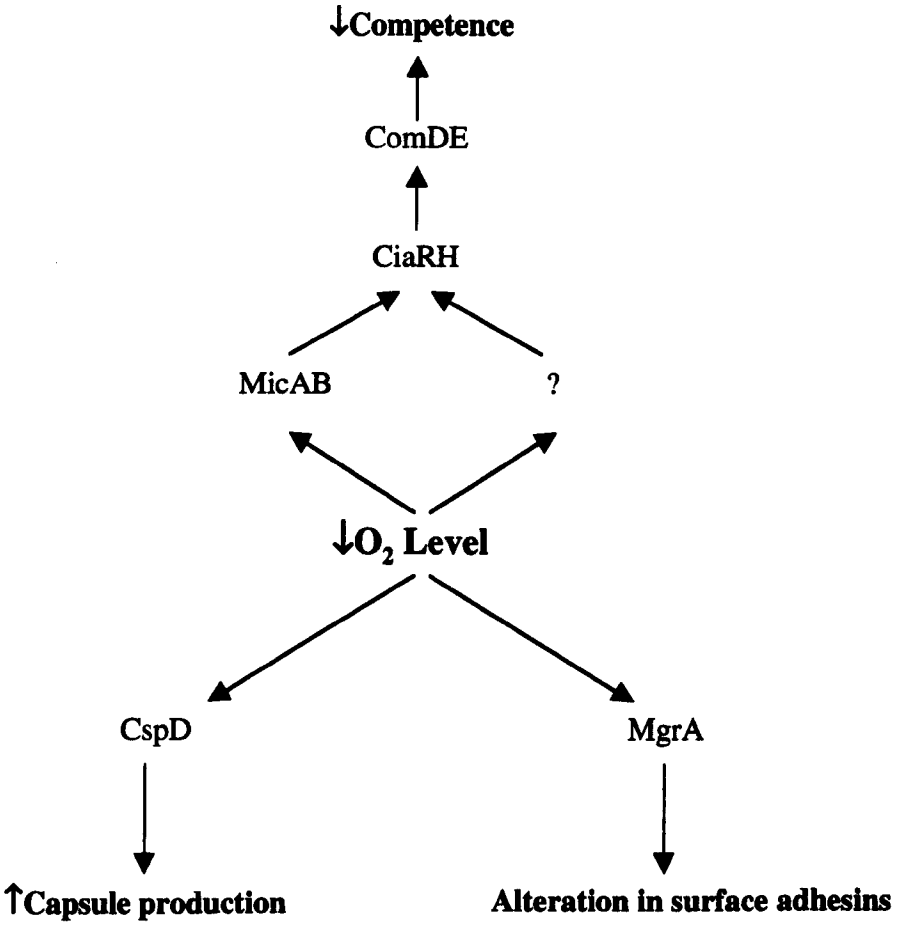


Figure 7.2 Model of the regulation of cellular processes by oxygen levels
Change in environmental oxygen levels regulates at least three different cellular processes in serotype 4 *S. pneumoniae*. Reduction in oxygen results in repression of competence, increase in capsule production and probable alteration in cell-surface adhesins. Repression of competence is mediated by three TCSTS, MicAB, CiaRH and ComDE. Whether they function in the same or parallel pathways is unclear. Alteration in capsule production by oxygen levels may involve CspD and surface adhesins, MgrA.

GAS and other streptococci regulate the expression of virulence factors in response to changes in oxygen and carbon dioxide levels (Nakamura *et al.*, 2004). Mga and RofA are known to be involved in this pathway in GAS. MgA is activated by growth in elevated carbon-dioxide (Caparon *et al.*, 1992; McIver *et al.*, 1995). RofA is affected by superoxide levels and oxygen partial pressure (Fogg and Caparon, 1997; Kreikemeyer *et al.*, 2003). In contrast Nra does not appear to be regulated by similar conditions to RofA (Kreikemeyer *et al.*, 2003).

Expression of the PPI2 gene *srtD* in *S. pneumoniae* is responsive to changes in oxygen and carbon dioxide levels and this requires the presence of MgrA (Figure 5.11). Decreases in both oxygen and carbon dioxide levels result in reduced *srtD* transcription, as measured by β-galactosidase activity in a *srtD:lacZ* fusion strain. However the β-galactosidase remains constant in all three atmospheres tested if the *mgrA* gene is inactivated in this strain background. This is consistent with either increased repression of transcription of *srtD* by MgrA with reduction in oxygen or carbon dioxide levels or lifting of transcriptional repression of *srtD* by MgrA with increased oxygen or carbon dioxide levels. It is not clear, however, whether a change in oxygen or a change in carbon dioxide levels exerts the strongest effect or at what level the response is achieved and therefore one can only speculate as to the physiological relevance of this effect in the host. It has not been formally demonstrated here that the other PPI2 genes, such as *rrgABC*, also show increased transcription in conditions of reduced oxygen and carbon dioxide levels or that increased *srtD* transcription results in altered surface protein expression but I hypothesise that this is the case. The data could be consistent with the model that adhesin production is greatest and CPS production least in the nasopharynx thereby favouring adherence to epithelial cells. As oxygen levels decrease

in the lower respiratory tract and in areas of pathology, CPS production increases and production some surface adhesins decreases resulting in a bacterial phenotype that favours avoidance of opsonophagocytosis and clearance by the host immune cells.

It is not known what, if any, factors are involved in signaling between gas concentration sensing and alteration of the activation status of MgrA. MgrA does not possess a PAS domain and hence is unlikely to sense changes in oxygen levels directly. No conclusions can be drawn about the involvement of RlrA in the response to alterations in gaseous compositions of the environment as the level of activity of the *rlrA:lacZ* fusions in experiments performed in order to examine this, was very low. It is possible that this is not an adequate method to assess *rlrA* transcription. RofA activity has been shown to be affected by oxygen levels in GAS, but neither it nor RlrA possess PAS domains. MicB, the HK component of a TCSTS in *S. pneumoniae*, is the only PAS domain-containing protein in *S. pneumoniae* (TIGR4 genome) (Echenique and Trombe, 2001a) and has been implicated in oxygen sensing in this pathogen. Competence is normally repressed under oxygen limitation (Echenique *et al.*, 2000; Echenique and Trombe, 2001a, b). Mutation of the PAS domain in MicB abolishes the kinase activity of the recombinant protein and allows the expression of competence under conditions of reduced oxygen. The oxygen-responsive alteration in *S. pneumoniae* surface structures could also require MicB.

7.7 CONCLUSION

The factors that determine the virulence of *S. pneumoniae* are very complex. The requirement for certain virulence factors varies according to the environment that the pathogen encounters. The distribution of some virulence genes and their regulation also

varies between *S. pneumoniae* strains. This is illustrated by the work described in chapters 4, 5 and 6 investigating the function, regulation, and distribution of the transcriptional regulators MgrA and RlrA, both of which are required for nasopharyngeal carriage and production of pneumonia in mice by serotype 4 *S. pneumoniae*. Both regulators control the transcription of genes of a newly described PAI, PPI2, yet the distribution of the two regulators varies among *S. pneumoniae* strains. The gene encoding RlrA is present in PPI2, which is carried by relatively few *S. pneumoniae* serotypes and the gene encoding MgrA is located on the chromosome outside PPI2 and appears to be present among most serotypes. It is likely that MgrA controls the transcription of additional virulence genes in serotypes that do not harbour PPI2 and thereby emphasizes the likelihood of a different repertoire of virulence genes among different serotypes and strains of this important human pathogen.

REFERENCES

(1997) Prevention of pneumococcal disease: recommendations of the advisory committee on immunisation practices. *MMWR Morb Mortal Wkly Rep* **46**: 1-24.

Abraham, E.P., Gardner, A.D., and Chain, E. (1941) Further observations on penicillin. *Lancet* **2**: 177-189.

Acebo, A., Nieto, C., Angeles Corrales, M., Espinosa, M., and Lopez, P. (2000a) Quantitative detection of *Streptococcus pneumoniae* cells harbouring single or multiple copies of the gene encodingthe green fluorescent protein. *Microbiology* **146**.

Acebo, P., Nieto, C., Corrales, M.A., Espinosa, M., and Lopez, P. (2000b) Quantitative detection of *Streptococcus pneumoniae* cells harbouring single or multiple copies of the gene encoding the green fluorescent protein. *Microbiology* **146**: 1267-1273.

Akerley, B.J., Cotter, P.A., and Miller, J.F. (1995) Ectopic expression of the flagellar regulon alters development of the Bordetella-host interaction. *Cell* **80**: 611-620.

Akerley, B.J., Rubin, E.J., Camilli, A., Lampe, D.J., Robertson, H.M., and Mekalanos, J.J. (1998) Systematic identification of essential genes by in vitro mariner mutagenesis. *Proc Natl Acad Sci U S A* **95**: 8927-8932.

AlonsoDeVelasco, E., Verheul, A.F., Verhoef, J., and Snippe, H. (1995) *Streptococcus pneumoniae*: virulence factors, pathogenesis, and vaccines. *Microbiol Rev* **59**: 591-603.

Angelichio, M.J., and Camilli, A. (2002) In vivo expression technology. *Infect Immun* **70**: 6518-6523.

Ashley Robinson, D., Briles, D., J., C.M., and Hollingshead, S. (2002) Evolution and virulence of serogroup 6 pneumococci on a global scale. *J Bacteriol* **184**: 6367-6375.

Austrian, R. (1977) Prevention of pneumococcal infection by immunization with capsular polysaccharides of *Streptococcus pneumoniae*: current status of polyvalent vaccines. *J Infect Dis* **136**: S38-42.

Austrian, R. (1986) Some aspects of the pneumococcal carrier state. *J Antimicrob Chemother* **18**: 35-45.

Autret, N., Dubail, I., Trieu-Cuot, P., Berche, P., and Charbit, A. (2001) Identification of new genes involved in the virulence of *Listeria monocytogenes* by signature-tagged transposon mutagenesis. *Infect Immun* **69**: 2054-2065.

Avery, O., MacLeod, C., and McCarty, M. (1944) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J Exp Med* **79**: 137-158.

Balachandran, P., Hollingshead, S.K., Paton, J.C., and Briles, D.E. (2001) The autolytic enzyme LytA of *Streptococcus pneumoniae* is not responsible for releasing pneumolysin. *J Bacteriol* **183**: 3108-3116.

Behr, M.A., Wilson, M.A., Gill, W.P., Salamon, H., Skolnik, G.K., Rane, S., and Small, P.M. (1999) Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**: 1520-1523.

Benton, K.A., Everson, M.P., and Briles, D.E. (1995) A pneumolysin-negative mutant of *Streptococcus pneumoniae* causes chronic bacteremia rather than acute sepsis in mice. *Infect Immun* **63**: 448-455.

Benton, K.A., Paton, J.C., and Briles, D.E. (1997) Differences in virulence for mice among Streptococcus pneumoniae strains of capsular types 2, 3, 4, 5, and 6 are not attributable to differences in pneumolysin production. *Infect Immun* **65**: 1237-1244.

Benton, K.A., VanCott, J.L., and Briles, D.E. (1998) Role of tumor necrosis factor alpha in the host response of mice to bacteremia caused by pneumolysin-deficient *Streptococcus pneumoniae*. *Infect Immun* **66**: 839-842.

Bergeron, Y., Ouellet, N., Deslauriers, A.M., Simard, M., Olivier, M., and Bergeron, M.G. (1998) Cytokine kinetics and other host factors in response to pneumococcal pneumonia in mice. *Infect Immun* **66**: 912-922.

Bergmann, S., Rohde, M., Chhatwal, G.S., and Hammerschmidt, S. (2001) alpha-Enolase of Streptococcus pneumoniae is a plasmin(ogen)-binding protein displayed on the bacterial cell surface. *Mol Microbiol* **40**: 1273-1287.

Bergmann, S., Rohde, M., and Hammerschmidt, S. (2004) Glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus pneumoniae* is a surface displayed plasminogen-binding protein. *Infect Immun* **72**: 2416-2419.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**: 9697-9702.

Berry, A.M., Lock, R.A., Hansman, D., and Paton, J.C. (1989) Contribution of autolysin to virulence of Streptococcus pneumoniae. *Infect Immun* **57**: 2324-2330.

Berry, A.M., Lock, R.A., Thomas, S.M., Rajan, D.P., Hansman, D., and Paton, J.C. (1994) Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant Escherichia coli. *Infect Immun* **62**: 1101-1108.

Berry, A.M., and Paton, J.C. (1996) Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of *Streptococcus pneumoniae*. *Infect Immun* **64**: 5255-5262.

Berry, A.M., and Paton, J.C. (2000) Additive attenuation of virulence of *Streptococcus pneumoniae* by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins. *Infect Immun* **68**: 133-140.

Beuzon, C.R., and Holden, D.W. (2001) Use of mixed infections with Salmonella strains to study virulence genes and their interactions in vivo. *Microbes Infect* **3**: 1345-1352.

Birrell, G.W., Brown, J.A., Wu, H.I., Giaever, G., Chu, A.M., Davis, R.W., and Brown, J.M. (2002) Transcriptional response of *Saccharomyces cerevisiae* to DNA-damaging agents does not identify the genes that protect against these agents. *Proc Natl Acad Sci U S A* **99**: 8778-8783.

Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J.R., Elvin, L., Ensor, K.M., Hackell, J., Siber, G., Malinoski, F., Madore, D., Chang, I., Kohberger, R., Watson, W., Austrian, R., and Edwards, K. (2000) Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Northern California Kaiser Permanente Vaccine Study Center Group. *Pediatr Infect Dis J* **19**: 187-195.

Blue, C.E., and Mitchell, T.J. (2003) Contribution of a response regulator to the virulence of *Streptococcus pneumoniae* is strain dependant. *Infect Immun* **71**: 4405-4413.

Bogart, D., de Groot, R., and Hermans, P.W.M. (2004) *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *Lancet Infect Dis* **4**: 144-154.

Boldrick, J.C., Alizadeh, A.A., Diehn, M., Dudoit, S., Liu, C.L., Belcher, C.E., Botstein, D., Staudt, L.M., Brown, P.O., and Relman, D.A. (2002) Stereotyped and specific gene expression programs in human innate immune responses to bacteria. *Proc Natl Acad Sci U S A* **99**: 972-977.

Boulnois, G.J., Paton, J.C., Mitchell, T.J., and Andrew, P.W. (1991a) Structure and function of pneumolysin, the multifunctional, thiol-activated toxin of Streptococcus pneumoniae. *Mol Microbiol* **5**: 2611-2616.

Boulnois, G.J., Paton, J.C., Mitchell, T.J., and Andrew, P.W. (1991b) Molecular analysis of pneymolysin, the multifunctional, thiol-activated toxin of Streptococcus pneumoniae. *Mol Microbiol* **5**: 2611-2616.

Bricker, A.L., and Camilli, A. (1999) Transformation of a type 4 encapsulated strain of *Streptococcus pneumoniae*. *FEMS Microbiol Lett* **172**: 131-135.

Briles, D.E., Hollingshead, S.K., Swiatlo, E., Brooks-Walter, A., Szalai, A., Virolainen, A., McDaniel, L.S., Benton, K.A., White, P., Prellner, K., Hermansson, A., Aerts, P.C., Van Dijk, H., and Crain, M.J. (1997) PspA and PspC: their potential for use as pneumococcal vaccines. *Microb Drug Resist* **3**: 401-408.

Brown, J., Gilliland, S.M., Spratt, B.G., and Holden, D.W. (2004) A locus within a variable region of pneumocccal pathogenicity island 1 contributes to virulence in mice. *Infect Immun* **72**: 1587-1593.

Brown, J.S., Gilliland, S.M., and Holden, D.W. (2001) A *Streptococcus pneumoniae* pathogenicity island encoding an ABC transporter involved in iron uptake and virulence. *Mol Microbiol* **40**: 572-585.

Brueggemann, A.B., Griffiths, D.T., Meats, E., Peto, T., Crook, D., and Spratt, B.G. (2003) Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *J Infect Dis* **187**: 1424-1432.

Butaye, P., Cloeckaert, A., and Schwarz, S. (2003) Mobile genes coding for efflux-mediated antimicrobial resistance in Gram-positive and Gram-negative bacteria. *Int J Antimicrob Agents* **22**: 205-210.

Camara, M., Boulnois, G.J., Andrew, P.W., and Mitchell, T.J. (1994) A neuraminidase from *Streptococcus pneumoniae* has the features of a surface protein. *Infect Immun* **62**: 3688-3695.

Camilli, A., Beattie, D.T., and Mekalanos, J.J. (1994) Use of genetic recombination as a reporter of gene expression. *Proc Natl Acad Sci U S A* **91**: 2634-2638.

Camilli, A., and Mekalanos, J.J. (1995) Use of recombinase gene fusions to identify Vibrio cholerae genes induced during infection. *Mol Microbiol* **18**: 671-683.

Camilli, A., Merrell, D.S., and Mekalanos, J.J. (2001) Strategies to Identify Bacterial Pathogenicity Factors. In *Principles of Bacterial Pathogenesis*. Groissman, E. (ed): Academic Press, pp. 133-177.

Campbell, E.A., Choi, S.Y., and Masure, H.R. (1998) A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol Microbiol* **27**: 929-939.

Canvin, J.R., Marvin, A.P., Sivakumaran, M., Paton, J.C., Boulnois, G.J., Andrew, P.W., and Mitchell, T.J. (1995) The role of pneumolysin and autolysin in the pathology of pneumonia and septicemia in mice infected with a type 2 pneumococcus. *J Infect Dis* **172**: 119-123.

Caparon, M.G., and Scott, J.R. (1987) Identification of a gene that regulates expression of M protein, the major virulence determinant of group A streptococci. *Proc Natl Acad Sci U S A* **84**: 8677-8681.

Caparon, M.G., Geist, R.T., Perez-Casal, J., and Scott, J.R. (1992) Environmental regulation of virulence in group A streptococci: transcription of the gene encoding M protein is stimulated by carbon dioxide. *J Bacteriol* **174**: 5693-5701.

Cauwels, A., Wan, E., Leismann, M., and Tuomanen, E. (1997) Coexistence of CD14-dependent and independent pathways for stimulation of human monocytes by gram-positive bacteria. *Infect Immun* **65**: 3255-3260.

Chapuy-Regaud, S., Ogunniyi, A.D., Diallo, N., Huet, Y., Desnottes, J.-F., Paton, J.C., Escaich, S., and Trombe, M.C. (2003) RegR, a glodal LacI/GalR family regulator, modulates virulence and competence in *Streptococcus pneumoniae*. *Infect Immun* **71**: 2615-2625.

Chastanet, A., Prudhomme, M., Claverys, J.P., and Msadek, T. (2001) Regulation of *Streptococcus pneumoniae clp* genes and their role in competence development and stress survival. *J Bacteriol* **183**: 7295-7307.

Chen, C., Bormann, N., and Clearly, P.P. (1993) ViR and Mry are homologous trans-acting regulators of M protein and C5a peptidase expression in group A streptococci. *Mol Gen Genet* **241**: 685-693.

Chhatwal, G.S. (2002) Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors. *Trends Microbiol* **10**: 205-208.

Chiang, S.L., Mekalanos, J.J., and Holden, D.W. (1999) In vivo genetic analysis of bacterial virulence. *Annu Rev Microbiol* **53**: 129-154.

Claverys, J.P., Dintilhac, A., Pestova, E.V., Martin, B., and Morrison, D.A. (1995) Construction and evaluation of new drug-resistance cassettes for gene disruption mutagenesis in Streptococcus pneumoniae, using an ami test platform. *Gene* **164**: 123-128.

Coffey, T.J., Dowson, C.G., Daniels, M., Zhou, J., Martin. C, Spratt, B.G., and Musser, J.M. (1991) Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *S. pneumoniae*. *Mol Microbiol* **5**: 2255-2260.

Coffey, T.J., Enright, M.C., Daniels, M., Morona, J.K., Morona, R., Hryniewicz, W., Paton, J.C., and Spratt, B.G. (1998) Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae. *Mol Microbiol* **27**: 73-83.

Coffey, T.J., Daniels, M., Enright, M.C., and Spratt, B.G. (1999) Serotype 14 variants of the Spanish penicillin-resistant serotype 9V clone of *Streptococcus pneumoniae* arose by large recombinational replacements of the cpsA-pbp1a region. *Microbiology* **145 (Pt 8)**: 2023-2031.

Crain, M.J., II, W.D.W., Turner, J.S., and Yother, J. (1990) Pnuemococcal surface protein A (PspA) is serologically highly varible and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. *Infect Immun* **58**: 3293-3299.

Cue, D., Lam, H., and Clearly, P.P. (2001) Genetic dissection of the *Streptococcus pyogenes* M1 protein: regions involved in fibronectin binding and intracellular invasion. *Microb Pathog* **31**: 231-242.

Cundell, D., Weiser, J.N., Shen, J., Young, A., and Tuomanen, E. (1995a) Relationship between colonial morphology and adherence of *Streptococcus pneumoniae*. *Infect Immun* **63**: 757-761.

Cundell, D.R., Gerard, N.P., Gerard, C., Idanpaan-Heikkila, I., and Tuomanen, E.I. (1995b) *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature* **377**: 435-438.

Cundell, D.R., Weiser, J.N., Shen, J., Young, A., and Tuomanen, E.I. (1995c) Relationship between colonial morphology and adherence of *Streptococcus pneumoniae*. *Infect Immun* **63**: 757-761.

Dagkessamanskaia, A., Moscoso, M., Henard, V., Guiral, S., Overweg, K., Reuter, M., Martin, B., Wells, J., and Claverys, J.P. (2004) Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of ciaR mutant cells. *Mol Microbiol* **51**: 1071-1086.

de Saizieu, A., Gardes, C., Flint, N., Wagner, C., Kamber, M., Mitchell, T.J., Keck, W., Amrein, K.E., and Lange, R. (2000) Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. *J Bacteriol* **182**: 4696-4703.

deLisa, M.P., Wu, C., Wang, L., Valdes, J.J., and Bentley, W.E. (2001) DNA microarray-based identification of genes controlled by autoinducer 2-stimulated quorum sensing in *Escherichia coli*. *J Bacteriol* **183**: 5239-5247.

Dintilhac, A., Alloing, G., Granadel, C., and Claverys, J.P. (1997) Competence and virulence of Streptococcus pneumoniae: Adc and PsaA mutants exhibit a requirement for Zn and Mn resulting from inactivation of putative ABC metal permeases. *Mol Microbiol* **25**: 727-739.

Doern, G.V., Heilmann, K.P., Huynh, H.K., Rhomberg, P.R., Coffman, S.L., and Brueggemann, A.B. (2001) Antimicrobial resistance among clinical isolates of Streptococcus pneumoniae in the United States during 1999--2000, including a comparison of resistance rates since 1994--1995. *Antimicrob Agents Chemother* **45**: 1721-1729.

Doliana, R., Bonaldo, P., and Colombatti, A. (1993) Type A modules: interacting domains founnd in several non-fibrillar collagens and in other extracellular matrix proteins. *Matrix* **13**: 297-306.

Echenique, J.R., Chapuy-Regaud, S., and Trombe, M.C. (2000) Competence regulation by oxygen in Streptococcus pneumoniae: involvement of ciaRH and comCDE. *Mol Microbiol* **36**: 688-696.

Echenique, J.R., and Trombe, M.C. (2001a) Competence repression under oxygen limitation through the two-component MicAB signal-transducing system in *Streptococcus pneumoniae* and involvement of the PAS domain of MicB. *J Bacteriol* **183**: 4599-4608.

Echenique, J.R., and Trombe, M.C. (2001b) Competence modulation by the NADH oxidase of Streptococcus pneumoniae involves signal transduction. *J Bacteriol* **183**: 768-772.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**: 14863-14868.

Enright, M.C., and Spratt, B.G. (1998) A multilocus sequence typing scheme for *Streptococcus pneumoniae:* identification of clones associated with seriuos invasive disease. *Microbiology* **144**: 3049-3060.

Eskola, J., Kilpi, T., Palmu, A., Jokinen, J., Haapakoski, J., Herva, E., Takala, A., Kayhty, H., Karma, P., Kohberger, R., Siber, G., and Makela, P.H. (2001) Efficacy of a pneumococcal conjugate vaccine against acute otitis media. *N Engl J Med* **344**: 403-409.

Feil, E.J., Enright, M.C., and Spratt, B.G. (2000) Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between Neisseria meningitidis and Streptococcus pneumoniae. *Res Microbiol* **151**: 465-469.

Feldman, C., Mitchell, T.J., Andrew, P.W., Boulnois, G.J., Read, R.C., Todd, H.C., Cole, P.J., and Wilson, R. (1990) The effect of *Streptococcus pneumoniae* pneumolysin on human respiratory epithelium in vitro. *Microb Pathog* **9**: 275-284.

Feldman, C., Reed, R., Rutmann, A., Munro, N.C., Jeffrey, D.K., Brian, A., Lund, V., Mitchell, T.J., Andrew, P.W., Boulnois, G.J., Todd, H.C., Cole, P.J., and Wilson, R. (1991) The effect of *Streptococcus pneumoniae* on inatct respiratory epithelium. *Eur Respir J* **5**: 576-583.

Filipe, S.R., Severina, E., and Tomasz, A. (2001) The role of murMN operon in penicillin resistance and antibiotic tolerance of Streptococcus pneumoniae. *Microb Drug Resist* **7**: 303-316.

Filipe, S.R., Severina, E., and Tomasz, A. (2002) The murMN operon: a functional link between antibiotic resistance and antibiotic tolerance in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **99**: 1550-1555.

Flanagan, P.G., and Paull, A. (1998) Carbon dioxide requirements of Streptococcus pneumoniae and Haemophilus influenzae. *J Antimicrob Chemother* **42**: 669-670.

Fogg, G.C., Gibson, C.M., and Caparon, M.G. (1994) The identification of rofA, a positive-acting regulatory component of prtF expression: use of an m gamma delta-based shuttle mutagenesis strategy in Streptococcus pyogenes. *Mol Microbiol* **11**: 671-684.

Fogg, G.C., and Caparon, M.G. (1997) Constitutive expression of fibronectin binding in Streptococcus pyogenes as a result of anaerobic activation of rofA. *J Bacteriol* **179**: 6172-6180.

Foster, T.J., and Hook, M. (1998) Surface protein adhesins of Staphylococcus aureus. *Trends Microbiol* **6**: 484-488.

Fox, M.S., and Allen, M.K. (1964) On the mechanism of deoxyribonucleate integration in pneumococcal transformation. *Proc Natl Acad Sci U S A* **52**: 412-419.

Gahan, C.G., and Hill, C. (2000) The use of listeriolysin to identify *in vivo* induced genes in the gram-positive intracellular pathogen *Listeria monocytogenes*. *Mol Microbiol* **36**: 498-507.

Garau, J. (2002) Treatment of drug-resistant pneumococcal pneumonia. *Lancet Infect Dis* **2**: 404-415.

Garcia, E., Llull, D., Munoz, R., Mollerach, M., and Lopez, R. (2000) Current trends in capsular polysaccharide biosynthesis of Streptococcus pneumoniae. *Res Microbiol* **151**: 429-435.

Garcia, J.L., Sanchez-Beato, A.R., Medrano, F.J., and Lopez, R. (1998) Versatility of choline-binding domain. *Microb Drug Resist* **4**: 25-36.

Gentry-Weeks, C., Coburn, P.S., and Gilmore, M.S. (2002) Phages and other mobile virulence elements in gram-positive pathogens. *Curr Top Microbiol Immunol* **264**: 79-94.

Ghaffar, F., Friedland, I.R., and McCracken, G.H., Jr. (1999) Dynamics of nasopharyngeal colonization by Streptococcus pneumoniae. *Pediatr Infect Dis J* **18**: 638-646.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., Arkin, A.P., Astromoff, A., El-Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.D., Flaherty, P., Foury, F., Garfinkel, D.J., Gerstein, M., Gotte, D., Guldener, U., Hegemann, J.H., Hempel, S., Herman, Z., Jaramillo, D.F., Kelly, D.E., Kelly, S.L., Kotter, P., LaBonte, D., Lamb, D.C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S.L., Revuelta, J.L., Roberts, C.J.R., M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D.D., Sookhai-Mahadeo, S., Storms, R.K., Strathern, J.N., Valle, G., Voet, M., Volckaert, G., Wang, C.Y., Ward, T.R., Wilhelmy, J., Winzeler, E.A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J.D., Snyder, M., Philippsen, P., Davis, R.W., and Johnston, M. (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**: 387-391.

Giammarinaro, P., and Paton, J.C. (2002) Role of RegM, a homologue of the catabolite repressor protein CcpA, in the virulence of *Streptococcus pneumoniae*. *Infect Immun* **70**: 5454-5461.

Gingles, N.A., Alexander, J.E., Kadioglu, A., Andrew, P.W., Kerr, A., Mitchell, T.J., Hopes, E., Denny, P., Brown, S., Jones, H.B., Little, S., Booth, G.C., and McPheat, W.L. (2001) Role of genetic resistance in invasive pneumococcal infection: identification and study of susceptibility and resistance in inbred mouse strains. *Infect Immun* **69**: 426-434.

Gollub, J., Ball, C.A., Binkley, G., Demeter, J., Finkelstein, D.B., Herbert, J.M., Hernandez-Boussard, T., Jin, H., Kaloper, M., Maltese, J.C., Schroeder, M., O., B.P., Botstein, D., and Sherlock, G. (2003) The Standford Microarray Database: data access and quality assessment tools. *Nucleic Acids Res* **31**: 94-96.

Gosink, K.K., Mann, E.R., Guglielmo, C., Tuomanen, E.I., and Masure, H.R. (2000) Role of novel choline binding proteins in virulence of Streptococcus pneumoniae. *Infect Immun* **68**: 5690-5695.

Gray, B.M., Converse, G.M., 3rd, and Dillon, H.C., Jr. (1980) Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. *J Infect Dis* **142**: 923-933.

Griffith, F. (1928) The significance of pneumococcal types. *J Hyg* **27**: 113-159.

Grosjean, H., and Fiers, W. (1982) Preferential codon usage in procaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* **18**: 199-209.

Guenzi, E., Gasc, A.M., Sicard, M.A., and Hakenbeck, R. (1994) A two-component signal-transducing system is involved in competence and penicillin susceptibility in laboratory mutants of Streptococcus pneumoniae. *Mol Microbiol* **12**: 505-515.

Hacker, J., Knapp, S., and Goebel, W. (1983) Spontaneous deletions and flanking regions of the chromosomal inherited hemolysin determinant of an *Escherichia coli* 06 strain. *J Bacteriol* **154**: 1145-1152.

Hacker, J., Blum-Oehler, G., Muhldorfer, I., and Tschape, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**: 1089-1097.

Hacker, J., and Kaper, J.B. (2000) Pathogenicity islands and evolution of microbes. *Annu Rev Microbiol* **54**: 641-679.

Hakenbeck, R., Grebe, T., Zahner, D., and Stock, J.B. (1999) Beta-lactam resistance in Streptococcus pneumoniae: penicillin-binding proteins and non-penicillin-binding proteins. *Mol Microbiol* **33**: 673-678.

Hakenbeck, R., Kaminski, K., Konig, A., van der Linden, M., Paik, J., Riechmann, P., and Zahner, D. (2000) Penicillin-Binding Proteins in β-Lactam-Resistant *Streptococcus pneumoniae*. In *Streptococcus pneumoniae Molecular Biology & Mechanisms of Disease*. Tomasz, A. (ed). New York, NY: Mary Ann Liebert, Inc., pp. 433-441.

Hakenbeck, R., Balmelle, N., Weber, B., Gardes, C., Keck, W., and de Saizieu, A. (2001) Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae. *Infect Immun* **69**: 2477-2486.

Hammerschmidt, S., Bethe, G., Remane, P.H., and Chhatwal, G.S. (1999) Identification of pneumococcal surface protein A as a lactoferrin- binding protein of Streptococcus pneumoniae. *Infect Immun* **67**: 1683-1687.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**: 557-580.

Harell, M., Mover-Lev, D., Levy, D., and Sade, J. (1996) Gas composition of the human nose and nasopharyngeal space. *Acta Otolaryngol* **116**: 82-84.

Hatfull, G.F., and Grindley, N.D.F. (1988) Resolvases and DNA-invertases; a family of enzymes active in site-specific recombination. In *Genetic Recombination*. Kucherlapati, R. and Smith, G.R. (eds). Washington D. C.: American Society for Microbiology, pp. 357-396.

Hausdorff, W.P., Bryant, J., Kloek, C., Paradiso, P.R., and Siber, G.R. (2000a) The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II. *Clin Infect Dis* **30**: 122-140.

Hausdorff, W.P., Bryant, J., Paradiso, P.R., and Siber, G.R. (2000b) Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* **30**: 100-121.

Hava, D.L., and Camilli, A. (2001) Isolation and characterization of a temperature-sensitive generalized transducing bacteriophage for Vibrio cholerae. *J Microbiol Methods* **46**: 217-225.

Hava, D.L., and Camilli, A. (2002) Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Molecular Microbiology* **45**: 1389-1405.

Hava, D.L., Hemsley, C.J., and Camilli, A. (2003a) Transcriptional Regulation in the *Streptococcus pneumoniae* rlrA Pathogenicity Islet by RlrA. *J Bacteriol* **185**: 413-421.

Hava, D.L., LeMieux, J., and Camilli, A. (2003b) From nose to lung: the regulation behind *Streptococcus pneumoniae* virulence factors. *Mol Microbiol* **50**: 1103-1110.

Havarstein, L.S., Coomaraswamy, G., and Morrison, D.A. (1995) An unmodified heptadecapeptide pheromone induces competence for genetic transformation in Streptococcus pneumoniae. *Proc Natl Acad Sci U S A* **92**: 11140-11144.

Havarstein, L.S., Hakenbeck, R., and Gaustad, P. (1997) Natural competence in the genus *Streptococcus*: evidence that streptococci can change pherotype by interspecies recombinational exchanges. *J Bacteriol* **179**: 6589-6594.

Heffron, F. (1983) In *Mobile Genetic Elements*. Shapiro, J. (ed). New York: Academic Press, pp. 223-260.

Hemsley, C., Joyce, E., Hava, D.L., Kawale, A., and Camilli, A. (2003) MgrA, an orthologue of Mga, Acts as a transcriptional repressor of the genes within the rlrA pathogenicity islet in Streptococcus pneumoniae. *J Bacteriol* **185**: 6640-6647.

Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**: 400-403.

Hergilis, L., and Magnuson, B. (1997) Middle ear composition in pathologic conditions: mass spectroscopy in otitis media with effusion and atelectasis. *Ann Otol Rhinol Laryngol* **106**: 743-745.

Holmes, A.R., McNab, R., Millsap, K.W., Rohde, M., Hammerschmidt, S., Mawdsley, J.L., and Jenkinson, H.F. (2001) The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence. *Mol Microbiol* **41**: 1395-1408.

Hoskins, J., Alborn, W.E., Jr., Arnold, J., Blaszczak, L.C., Burgett, S., DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A.R., Lagace, R.E., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S.M., McHenney, M., McLeaster, K., Mundy, C.W., Nicas, T.I., Norris, F.H., O'Gara, M., Peery, R.B., Robertson, G.T., Rockey, P., Sun, P.M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C.A., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R., Jr., Skatrud, P.L., and Glass, J.I. (2001) Genome of the bacterium Streptococcus pneumoniae strain R6. *J Bacteriol* **183**: 5709-5717.

Hotchkiss, R.D. (1951) Transfer of penicillin resistance in pneumococci by the deoxyridonucleate derived from resistant cultures. *Cold Spring Harbour Symp. Quant. Biol.* **16**: 457-460.

Huebner, R.E., Wasas, A.D., and Klugman, K.P. (2000) Trends in antimicrobial resistance and serotype distribution of blood and cerebrospinal fluid isolates of Streptococcus pneumoniae in South Africa, 1991-1998. *Int J Infect Dis* **4**: 214-218.

Hui, F.M., Zhou, L., and Morrison, D.A. (1995) Competence for genetic transformation in *S. pneumoniae*: organisation of a regulatory locus with homology to lactococcin A secretion genes. *Gene* **153**: 25-31.

Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J Mol Biol* **146**: 1-21.

Isberg, R.R., and Leong, R.R. (1990) Multiple beta-1 chain integrins are receptors for invasin, a protein that promotes bacterial penetration into mammalian cells. *Cell* **60**: 861-871.

Jaffe, J., Natanson-Yaron, S., Caparon, M., and Hanski, E. (1996) Abstract Protein F2, a novel fibronectin-binding protein from Streptococcus pyogenes, possesses two binding domains. *Mol Microbiol* **21**: 373-384.

Jedrzejas, M.J. (2001) Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev* **65**: 187-207 ; first page, table of contents.

Joh, D., Wann, E.R., Kreikemeyer, B., Speziale, P., and Hook, M. (1999) Role of fibronectin-binding MSCRAMMs in bacterial adherence and entry into mammalian cells. *Matrix Biol* **18**: 211-223.

Johnson, M.K. (1977) Cellular location of pneumolysin. *FEMS Microbiol Lett* **2**: 243-245.

Jones, A.L., Knoll, K.M., and Rubens, C.E. (2000) Identification of Streptococcus agalactiae virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. *Mol Microbiol* **37**: 1444-1455.

Kadioglu, A., Gingles, N.A., Grattan, K., Kerr, A., Mitchell, T.J., and Andrew, P.W. (2000) Host cellular immune response to pneumococcal lung infection in mice. *Infect Immun* **68**: 492-501.

Kadioglu, A., Taylor, S., Iannelli, F., Pozzi, G., Mitchell, T.J., and Andrew, P.W. (2002) Upper and lower respiratory tract infection by Streptococcus pneumoniae is affected by pneumolysin deficiency and differences in capsule type. *Infect Immun* **70**: 2886-2890.

Kadioglu, A., Echenique, J.R., Manco, S., Trombe, M.C., and Andrew, P.W. (2003) The MicAB two-component signalling system is involved in virulence of *Streptococcus pneumoniae*. *Infect Immun* **71**: 6676-6679.

Kamerling, J. (2000) Pneumococcal Polysaccharides: A Chemical View. In *Streptococcus pneumoniae - Molecular Biology & Mechanisms of Disease.* Tomasz, A. (ed). Larchmont, NY: Mary Ann Lieber, Inc., pp. 81-114.

Kaper, J.B., and Hacker, J. (2000) Pathogenicity Islands and the evolution of Microbes. *Annu Rev Microbiol* **54**: 641-479.

Kato-Maeda, M., Gao, Q., and Small, P.M. (2001) Microarray analysis of pathogens and their interaction with hosts. *Cell Microbiol* **3**: 713-719.

Kharat, A.S., and Tomasz, A. (2003) Inactivation of the *srtA* gene affects localisation of surface proteins and decreases adhesion of *Streptococcus pneumoniae* to human pharyngeal cells *in vitro*. *Infect Immun* **71**: 2758-2765.

Kihlberg, B.M., Cooney, J., Caparon, M.G., Olsen, A., and Bjorck, L. (1995) Biological properties of a Streptococcus pyogenes mutant generated by Tn916 insertion in mga. *Microb Pathog* **19**: 299-315.

Kilpi, T., Ahman, H., Jokinen, J., Lankinen, K.S., Palmu, A., Savolainen, H., Gronholm, M., Leinonen, M., Hovi, T., Eskola, J., Kayhty, H., Bohidar, N., Sadoff, J.C., Makela, P.H., and Group., F.O.M.S. (2003) Protective efficacy of a second pneumococcal conjugate vaccine against pneumococcal acute otitis media in infants and children: randomized, controlled trial of a 7-valent pneumococcal polysaccharide-meningococcal outer membrane protein complex conjugate vaccine in 1666 children. *Clin Infect Dis* **37**: 1155-1164.

Kim, J.O., and Weiser, J.N. (1998) Association of intrastrain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of Streptococcus pneumoniae. *J Infect Dis* **177**: 368-377.

Klugman, K. (2001a) Efficacy of pneumococcal conjugate vaccines and their effect on carriage and antimicrobial resistance. *Lancet Infect Dis* **1**: 85-91.

Klugman, K.P. (2001b) Efficacy of pneumococcal conjugate vaccines and their effect on carriage and antimicrobial resistance. *Lancet Infect Dis* **1**: 85-91.

Kolter, R., Inuzuka, M., and Helinski, D.R. (1978) Trans-complementation-dependent replication of a low molecular weight origin fragment from plasmid R6K. *Cell* **15**: 1199-1208.

Kononen, E., Jousimies-Somer, H., Bryk, A., Kilp, T., and Kilian, M. (2002) Establishment of streptococci in the upper respiratory tract: longitudinal changes in the mouth and nasopharynx up to 2 years of age. *J Med Microbiol* **51**: 723-730.

Kreikemeyer, B., McIver, K., and Pobielski, A. (2003) Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* **11**: 224-232.

Krivan, H.C., Roberts, D.D., and Ginsburg, V. (1988) Many pulmonary pathogenic bacteria bind specifically to the carbohydrate sequence GalNAc beta 1-4Gal found in some glycolipids. *Proc Natl Acad Sci U S A* **85**: 6157-6161.

Kuhar, I., Putten, J.P., Zgur-Bertok, D., Gaastra, W., and Jordi, B.J. (2001) Codon-usage based regulation of colicin K synthesis by stress alarmone ppGpp. *Mol Microbiol* **41**: 207-216.

Lacks, S.A., and Greenberg, B. (1976) Single-strand breakage on binding of DNA to cells in the genetic transformation of *Diplococcus pneumoniae*. *J Mol Biol* **101**: 255-275.

Lampe, D.J., Churchill, M.E., and Robertson, H.M. (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *Embo J* **15**: 5470-5479.

Lange, R., Wagner, C., de Saizieu, A., Flint, N., Molnos, J., Stieger, M., Caspers, P., Kamber, M., Keck, W., and Amrein, K.E. (1999) Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*. *Gene* **237**: 223-234.

Lau, G.W., Haataja, S., Lonetto, M., Kensit, S.E., Marra, A., Bryant, A.P., McDevitt, D., Morrison, D.A., and Holden, D.W. (2001) A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. *Mol Microbiol* **40**: 555-571.

Lee, M.S., and Morrison, D.A. (1999) Identification of a new regulator in Streptococcus pneumoniae linking quorum sensing to competence for genetic transformation. *J Bacteriol* **181**: 5004-5016.

Lee, S.H., Angelichio, M.J., Mekalanos, J.J., and Camilli, A. (1998) Nucleotide sequence and spatiotemporal expression of the Vibrio cholerae vieSAB genes during infection. *J Bacteriol* **180**: 2298-2305.

Lee, S.H., Hava, D.L., Waldor, M.K., and Camilli, A. (1999) Regulation and temporal expression patterns of Vibrio cholerae virulence genes during infection. *Cell* **99**: 625-634.

Lee, S.H., Butler, S.M., and Camilli, A. (2001) Selection for in vivo regulators of bacterial virulence. *Proc Natl Acad Sci U S A* **98**: 6889-6894.

Leino, T., Auranen, K., Jokinen, J., Leinonen, M., Tervonen, P., and Takala, A. (2001) Pneumococcal carriage in children during their first two years: important role of family exposure. *Paediatr Infect Dis* **20**: 1022-1027.

Li, J., Swanson, R.V., Simon, M.I., and Weis, R.M. (1995) The response regulators CheB and CheY exhibit competitive binding to the kinase CheA. *Biochemistry* **34**: 14626-14636.

Lindsay, J.A., Ruzin, A., Ross, H.F., Kurepina, N., and Novick, R.P. (2001) Molecular genetics of SaPI1--a mobile pathogenicity island in Staphylococcus aureus. *Mol Microbiol* **41**: 365-377.

Lipsitch, M., and Moxon, E.R. (1997) Virulence and transmissibility of pathogens: what is the relationship? *Trends Microbiol* **5**: 31-37.

Lottenberg, R., Minning Wenz, D., and Boyle, M.D. (1994) Capturing host plasmin(ogen): a common mechanism for invasive pathogens? *Trends Microbiol* **2**: 20-24.

Lottenberg, R. (1997) A novel approach to explore the role of plasminogen in bacterial pathogenesis. *Trends Microbiol* **5**: 466-467.

Lowe, A.M., Beattie, D.T., and Deresiewicz, R.L. (1998) Identification of novel staphylococcal virulence genes by in vivo expression technology. *Mol Microbiol* **27**: 967-976.

Lu, L., Lamm, M.E., Li, H., Corthesy, B., and Zhang, J.R. (2003) The human polymeric immunoglobulin receptor binds to *Streptococcus pneumoniae* via domains 3 and 4. *J Biol Chem* **278**: 48178-48187.

Luo, P., and Morrison, D.A. (2003) Transient association of an alternative sigma factor, ComX, with RNA polymerase during the period of competence for genetic transformation in *Streptococcus pneumoniae*. *J Bacteriol* **185**: 349-358.

Magee, A.D., and Yother, J. (2001) Requirement for capsule in colonization by Streptococcus pneumoniae. *Infect Immun* **69**: 3755-3761.

Mahan, M.J., Slauch, J.M., Hanna, P.C., Camilli, A., Tobias, J.W., Waldor, M.K., and Mekalanos, J.J. (1993) Selection for bacterial genes that are specifically induced in host tissues: the hunt for virulence factors. *Infect Agents Dis* **2**: 263-268.

Mahan, M.J., Tobias, J.W., Slauch, J.M., Hanna, P.C., Collier, R.J., and Mekalanos, J.J. (1995) Antibiotic-based selection for bacterial genes that are specifically induced during infection of a host. *Proc Natl Acad Sci U S A* **92**: 669-673.

Malley, R., Lipsitch, M., Stack, A., Saladino, R., Fleisher, G., Pelton, S., Thompson, C., Briles, D., and Anderson, P. (2001) Intranasal immunization with killed unencapsulated whole cells prevents colonization and invasive disease by capsulated pneumococci. *Infect Immun* **69**: 4870-4873.

Mangtani, P., Cutts, F., and Hall, A.J. (2003) Efficacy of polysaccharide vaccine in adults in more developed countries: the stae of the evidence. *Lancet Infect Dis* **3**: 71-78.

Marra, A., Asundi, J., Bartilson, M., Lawson, S., Fang, F., Christine, J., Wiesner, C., Brigham, D., Schneider, W.P., and Hromockyj, A.E. (2002a) Differential fluorescence induction analysis of Streptococcus pneumoniae identifies genes involved in pathogenesis. *Infect Immun* **70**: 1422-1433.

Marra, A., Lawson, S., Asundi, J.S., Brigham, D., and Hromockyj, A.E. (2002b) In vivo characterization of the psa genes from Streptococcus pneumoniae in multiple models of infection. *Microbiology* **148**: 1483-1491.

Martin, B., Prudhomme, M., Alloing, G., Granadel, C., and Claverys, J.P. (2000) Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. *Mol Microbiol* **38**: 867-878.

Massey, R.C., Kantzanou, M.N., Fowler, T., Day, N.P., Schofield, K., Wann, E.R., Berendt, A.R., Hook, M., and Peacock, S.J. (2001) Fibronectin-binding protein A of *Staphylococcus aureus* has multiple, substituting, binding regions that mediate adherence to fibronectin and invasion of endothelial cells. *Cell Microbiol* **3**: 839-851.

Mazmanian, S.K., Liu, G., Ton-That, H., and Schneewind, O. (1999) Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall. *Science* **285**: 760-763.

Mazmanian, S.K., Ton-That, H., and Schneewind, O. (2001) Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Mol Microbiol* **40**: 1049-1057.

Mbelle, M., Huebner, R., Wasas, A.D., Kimura, A., Chang, I., and Klugman, K. (1999) Immunogenicity and effect on nsaopharyngeal carriage of a nonavalent pneumococcal vaccine. *J Infect Dis* **180**: 1171-1176.

McCluskey, J., Hinds, J., Husain, S., Witney, A., and Mitchell, T.J. (2004) A two-component system that controls the expression of pneumococccal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. *Molecular Microbiology* **51**: 1661-1675.

McCullers, J.A., and Tuomanen, E.I. (2001a) Molecular pathogenesis of pneumococcal pneumonia. *Front Biosci* **6**: D877-889.

McCullers, J.A., and Tuomanen, E.I. (2001b) Molecular pathogenesis of pneumococcal pneumonia. *Front Biosci* **6**: D877-889.

McDaniel, L.S., Yother, J., Vijayakumar, M., McGarry, L., Guild, W.R., and Briles, D.E. (1987) Use of insertional inactivation to facilitate studies of biological properties of pneumococcal surface protein A (PspA). *J Exp Med* **165**: 381-394.

McGee, L., McDougal, L., Zhou, J., Spratt, B.G., Tenover, F.C., George, R., Hakenbeck, R., Hryniewicz, W., Lefevre, J.C., Tomasz, A., and Klugman, K.P. (2001) Nomenclature of major antimicrobial-resistant clones of Streptococcus pneumoniae defined by the pneumococcal molecular epidemiology network. *J Clin Microbiol* **39**: 2565-2571.

McIver, K., Heath, A.S., and Scott, J.R. (1995) Regulation of virulence by environmental signals in group A streptococci: Influence of osmolarity, temperature, Gas exchange and iron limitation on *emm* transcription. *Infect Immun* **63**: 4540-4542.

McIver, K., and Scott, J.R. (1997) Role of *mga* in growth phase regulation of Virulence genes of the group A streptococcus. *J Bacteriol* **179**: 5178-5187.

McIver, K., and Myles, R.L. (2002) Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. *Mol Microbiol* **43**: 1591-1601.

McLandsborough, L.A., and Cleary, P.P. (1995) Insertional inactivation of virR in Streptococcus pyogenes M49 demonstrates that VirR functions as a positive regulator of ScpA, FcRA, OF, and M protein. *FEMS Microbiol Lett* **128**: 45-51.

Meats, E., Brueggemann, A.B., Enright, M.C., Sleeman, K., Griffiths, D., Crook, D., and Spratt, B.G. (2003) Stability of serotypes during nasopharyngeal carriage of *Streptococcus pneumoniae*. *J Clin Microbiol* **41**: 386-392.

Mecsas, J. (2002) Use of signature-tagged mutagenesis in pathogenesis studies. *Curr Opin Microbiol* **5**: 33-37.

Mei, J.M., Nourbakhsh, F., Ford, C.W., and Holden, D.W. (1997) Identification of Staphylococcus aureus virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. *Mol Microbiol* **26**: 399-407.

Mengaud, J., Ohayon, H., Gounon, P., Mege, R.-M., and Cossart, P. (1996) E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells. *Cell* **84**: 923-932.

Merrell, D.S., and Camilli, A. (2000) Regulation of vibrio cholerae genes required for acid tolerance by a member of the "ToxR-like" family of transcriptional regulators. *J Bacteriol* **182**: 5342-5350.

Merrell, D.S., Butler, S.M., Qadri, F., Dolganov, N.A., Alam, A., Cohen, M.B., Calderwood, S.B., Schoolnik, G.K., and Camilli, A. (2002) Host-induced epidemic spread of the cholera bacterium. *Nature* **417**: 642-645.

Merrell, D.S., Goodrich, M.L., Otto, G., Tompkins, L.S., and Falkow, S. (2003) pH-regulated gene expression of te gastric pathogen *Helicobacter pylori*. *Infect Immun* **71**: 3529-3539.

Miller, J.H. (1972) Assay of Beta-galactosidase. In *Experiments in Molecular Genetics*. Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press, pp. 352-355.
Mitchell, T.J., Andrew, P.W., Saunders, F.K., Smith, A.N., and Boulnois, G.J. (1991) Complement activation and antibody binding by pneumolysin via a region of the toxin homologous to a human acute-phase protein. *Mol Microbiol* **5**: 1883-1888.
Mitchell, T.J., and Andrew, P.W. (1997) Biological properties of pneumolysin. *Microb Drug Resist* **3**: 19-26.
Mitchell, T.J. (2000) Virulence factors and the pathogenesis of disease caused by *Streptococcus pneumoniae*. *Res Microbiol* **151**: 413-419.
Molinari, G., Talay, S.R., Valentin-Weigand, P., Rohde, M., and Chhatwal, G.S. (1997) The fibronectin binding protein a of *Streptococcus pyogenes*, Sfb1, is involved in the internalisation of group A streptococci by epithelial cells. *Infect Immun* **65**: 1357-1363.
Morrison, D.A., and Lee, M.S. (2000) Regulation of competence for genetic transformation in Streptococcus pneumoniae: a link between quorum sensing and DNA processing genes. *Res Microbiol* **151**: 445-451.
Mortier-Barriere, I., de Saizieu, A., Claverys, J.P., and Martin, B. (1998) Competence-specific induction of recA is required for full recombination proficiency during transformation in Streptococcus pneumoniae. *Mol Microbiol* **27**: 159-170.
Msadek, T. (1999) When the going gets tough: survival strategies and environmental signaling networks in Bacillus subtilis. *Trends Microbiol* **7**: 201-207.
Nakamura, T., Hasegawa, T., Torii, K., Hasegawa, Y., Shimokata, K., and Ohta, M. (2004) Two-dimensional gel electrophoresis analysis of the abundance of virulent exoproteins of group A streptococcus caused by environmental changes. *Arch Microbiol* **181**: 74-81.
Nieto, C., Espinosa, M., and Puyets, A. (1997) The maltose/maltodextrin regulaon of Streptococcus pneumoniae. *J Biol Chem* **272**: 30860-30865.
Nieto, C., de Palencia, P.F., Lopez, P., and Espinosa, M. (2000) Construction of a tightly regulated plasmid vector for Streptococcus pneumoniae: Controlled expression of the green fluorescent protein. *Plasmid* **43**: 205-213.
Novak, R., Cauwels, A., Charpentier, E., and Tuomanen, E. (1999a) Identification of a Streptococcus pneumoniae gene locus encoding proteins of an ABC phosphate transporter and a two-component regulatory system. *J Bacteriol* **181**: 1126-1133.
Novak, R., Henriques, B., Charpentier, E., Normark, S., and Tuomanen, E. (1999b) Emergence of vancomycin tolerance in Streptococcus pneumoniae. *Nature* **399**: 590-593.
Novak, R., Charpentier, E., Braun, J.S., and Tuomanen, E. (2000) Signal transduction by a death signal peptide: uncovering the mechanism of bacterial killing by penicillin. *Mol Cell* **5**: 49-57.
O'Brien, K.L., Moulton, L., Reid, R., Weatherholtz, R., Oski, J., Brown, L., Kumar, G., Parkinson, A., Hu, D., Hackell, J., Chang, I., Kohberger, R., Siber, G., and Santosham, M. (2003) Efficacy and safety of seven-valent conjugate pneumococcal vaccine in American Indian children: group randomised trial. *Lancet* **362**: 355-361.
Obaro, S., Adegbola, R.A., Banya, W., and Greenwood, B. (1996) Carriage of pneumococci after pneumococcal vaccination. *Lancet* **348**: 271-272.
Obaro, S., and Aaberge, I.S. (2002) the pneumococcus: cariage disease and conjugate vaccines. *J Med Microbiol* **51**: 98-104.

Ogunniyi, A.D., Giammarinaro, P., and Paton, J.C. (2002) The genes encoding virulence-associated proteins and the capsule of *Streptococcus pneumoniae* are up-regulated and differentially expressed *in vivo*. *Microbiology* **148**: 2045-2053.

Orihuela, C.J., Janssen, R., Robb, C.W., Watson, D.A., and Niesel, D.W. (2000) Peritoneal culture alters *Streptococcus pneumoniae* protein profiles and virulence properties. *Infect Immun* **68**: 6082-6086.

Orihuela, C.J., Mills, J., Robb, C.W., Wilson, C.J., Watson, D.A., and Niesel, D.W. (2001) *Streptococcus pneumoniae* PstS production is phosphate responsive and enhanced during growth in the murine peritoneal cavity. *Infect Immun* **69**: 7565-7571.

Ottolenghi, E., and Hotchkiss, R.D. (1962) Release of genetic transforming agent from pneumococcal cultures during growth and disintegration. *J Exp Med* **116**: 491-519.

Ottolenghi, E., and MacLeod, C.M. (1963) Genetic transformation amongst living pneumococci in the mouse. *Proc Natl Acad Sci U S A* **50**: 491-519.

Ottolenghi-Nightingale, E. (1969) Spontaneously occurring bacterial transformations in mice. *J Bacteriol* **100**: 445-452.

Paton, J.C., and Ferrante, A. (1983) Inhibition of human polymorphonuclear leukocyte respiratory burst, bactericidal activity, and migration by pneumolysin. *Infect Immun* **41**: 1212-1216.

Paton, J.C., Rowan-Kelly, B., and Ferrante, A. (1984) Activation of human complement by the pneumococcal toxin pneumolysin. *Infect Immun* **43**: 1085-1087.

Paton, J.C., Andrew, P.W., Boulnois, G.J., and Mitchell, T.J. (1993) Molecular analysis of the pathogenicity of Streptococcus pneumoniae: the role of pneumococcal proteins. *Annu Rev Microbiol* **47**: 89-115.

Paton, J.C. (1996) The contribution of pneumolysin to the pathogenicity of Streptococcus pneumoniae. *T I M S* **4**: 103-106.

Patti, J.M., Allen, B.L., McGavin, M.J., and Hook, M. (1994) MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu Rev Microbiol* **48**: 585-617.

Pestova, E.V., and Morrison, D.A. (1998) Isolation and characterization of three Streptococcus pneumoniae transformation-specific loci by use of a lacZ reporter insertion vector. *J Bacteriol* **180**: 2701-2710.

Peterson, S., Cline, R.T., Tettelin, H., Sharov, V., and Morrison, D.A. (2000) Gene expression analysis of the *Streptococcus pneumoniae* competence regulons by use of DNA microarrays. *J Bacteriol* **182**: 6192-6202.

Peterson, S.N., Sung, C.K., Cline, R., Desai, B.V., Snesrud, E.C., Luo, P., Walling, J., Li, H., Mintz, M., Tsegaye, G., Burr, P.C., Do, Y., Ahn, S., Gilbert, J.V., Fleischmann, R.D., and Morrison, D.A. (2004) Identification of competence pheromone responsive genes in Streptococcus pneumoniae by use of DNA microarrays. *Mol Microbiol* **51**: 1051-1070.

Pobielski, A., Woischnik, M., Pohl, B., and Schmidt, K.H. (1996) What is the size of the group A streptococcal *vir* regulon? The Mga regulator affects expression of secreted and surface virulence factors. *Med Microbiol Immunol* **185**: 171-181.

Pobielski, A., Woischnik, M., Leonard, B.A., and Schmidt, K.H. (1999) Characterisation of *nra*, a global negative regulator gene in group A streptococci. *Molecular Microbiology* **31**: 1051-1064.

Podbielski, A., Woischnik, M., Leonard, B.A., and Schmidt, K.H. (1999) Characterization of nra, a global negative regulator gene in group A streptococci. *Mol Microbiol* **31**: 1051-1064.

Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L., and Simon, D. (1998) Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infect Immun* **66**: 5620-5629.

Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H., and Dennis, P.P. (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase beta subunit in *Escherichia coli*. *Proc Natl Acad Sci U S A* **76**: 1697-1701.

Prudhomme, M., Libante, V., and Claverys, J.P. (2002) Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **99**: 2100-2105.

Puyet, A., and Espinosa, M. (1993) Structure of the maltodextrin-uptake locus of *Streptococcus pneumoniae*. Correlation to the Escherichia coli maltose regulon. *J Mol Biol* **230**: 800-811.

Ramirez, M., Morrison, D.A., and Tomasz, A. (1997) Ubiquitous distribution of the competence related genes comA and comC among isolates of Streptococcus pneumoniae. *Microb Drug Resist* **3**: 39-52.

Rasmussen, J.M., Eden, A., and Bjorck, L. (2000) SclA, a novel collagen-like surface protein of *Streptococcus pyogenes*. *Infection and Immunity* **68**.

Rayner, C.F., Jackson, A.D., Rutman, A., Dewar, A., Mitchell, T.J., Andrew, P.W., Cole, P.J., and Wilson, R. (1995) Interaction of pneumolysin-sufficient and -deficient isogenic variants of *Streptococcus pneumoniae* with human respiratory mucosa. *Infect Immun* **63**: 442-447.

Ren, B., Szalai, A.J., Hollingshead, S.K., and Briles, D.E. (2004) Effects of PspA and antibodies to PspA on activation and deposition of complement on the pneumococcal surface. *Infect Immun* **72**: 114-122.

Rijneveld, A.W., Weijer, S., Florquin, S., Speelman, P., Shimizu, T., Ishii, S., and van der Poll, T. (2003) Improved host defense against pneumococcal pneumonia in platelet-activating factor receptor-deficient mice. *J Infect Dis* **189**: 711-716.

Rimini, R., Jansson, B., Feger, G., Roberts, T.C., de Francesco, M., Gozzi, A., Faggioni, F., Domenici, E., Wallace, D.M., Frandsen, N., and Polissi, A. (2000) Global analysis of transcription kinetics during competence development in Streptococcus pneumoniae using high density DNA arrays. *Mol Microbiol* **36**: 1279-1292.

Ring, A., Weiser, J.N., and Tuomanen, E.I. (1998) Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. *J Clin Invest* **102**: 347-360.

Robertson, G.T., Zhao, J., Desai, B.V., Coleman, W.H., Nicas, T.I., Gilmour, R., Grinius, L., Morrison, D.A., and Winkler, M.E. (2002) Vancomycin tolerance induced by erythromycin but not loss of *vncRS*, *vex3*, or *pep27* function in *Streptococcus pneumoniae*. *J Bacteriol* **184**: 6987-7000.

Robinson, M., Lilley, R., Emtage, J.S., Yarranton, G., Stephens, P., Millican, A., Eaton, M., and Humphreys, G. (1984) Codon uasge can effect translation of genes in *Escherichia coli*. *Nucleic Acids Res* **12**: 6663-6671.

Rogers, P.D., Thornton, J., Barker, K.S., McDaniel, D.O., Sacks, G.S., Swiatlo, E., and McDaniel, L.S. (2003) Pneumolysin-dependent and -independent gene expression identified by cDNA microarray analysis of THP-1 human mononuclear cells stimulated by*Streptococcus pneumoniae*. *Infect Immun* **71**: 2087-2094.

Ronda, C., Garcia, J.J., Garcia, C., Sanchez-Puelles, J.M., and Lopez, R. (1987) Biological role of pneumococcal amidase. *Eur J Biochem* **1987**: 621-624.

Rosenow, C., Ryan, P., Weiser, J.N., Johnson, S., Fontan, P., Ortqvist, A., and Masure, H.R. (1997) Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae. *Mol Microbiol* **25**: 819-829.

Rozdzinski, E., Marre, R., Susa, M., Wirth, R., and Muscholl-Silberhorn, A. (2001) Aggregation substance-mediated adherence of Enterococcus faecalis to immobilized extracellular matrix proteins. *Microb Pathog* **30**: 211-220.

Rubins, J.B., Charboneau, D., Paton, J.C., Mitchell, T.J., Andrew, P.W., and Janoff, E.N. (1995) Dual function of pneumolysin in the early pathogenesis of murine pneumococcal pneumonia. *J Clin Invest* **95**: 142-150.

Rubins, J.B., Charboneau, D., Fasching, C., Berry, A.M., Paton, J.C., Alexander, J.E., Andrew, P.W., Mitchell, T.J., and Janoff, E.N. (1996) Distinct roles for pneumolysin's cytotoxic and complement activities in the pathogenesis of pneumococcal pneumonia. *Am J Respir Crit Care Med* **153**: 1339-1346.

Rubins, J.B., and Janoff, E.N. (1998) Pneumolysin: a multifunctional pneumococcal virulence factor. *J Lab Clin Med* **131**: 21-27.

Sade, J., Luntz, M., and Levy, D. (1995) Middle ear gas composition and middle ear aeration. *Ann Otol Rhinol Laryngol* **104**: 369-373.

Sakai, H., Imamura, C., Osado, Y., Saito, R., Washio, T., and Tomita, M. (2001) Correlation between Shine-Delgarno sequence conservation and codon usage of bacterial genes. *J mol Evol* **52**: 164-170.

Salama, N., Guillemin, K., McDaniel, T.K., Sherlock, G., Tompkins, L., and Falkow, S. (2000) A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. *Proc Natl Acad Sci U S A* **97**: 14668-14673.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1998) *Molecular Cloning: A Laboratory Manual, Second Edition.* Plainview, NY: Cold Spring Harbor Lab. Press.

Sampson, J.S., O'Connor, S.P., Stinson, A.R., Tharpe, J.A., and Russell, H. (1994) Cloning and nucleotide sequence analysis of *psaA*, the *Streptococcus pneumoniae* gene encoding a 37-kilodalton protein hologous tp previously reported *Streptococcus sp.* adhesins. *Infect Immun* **62**: 319-324.

Sarkari, J., Pandit, N., Moxon, E.R., and Achtman, M. (1994) Variable expression of the Opc outer membrane protein in Neisseria meningitidis is caused by size variation of a promoter containing poly-cytidine. *Mol Microbiol* **13**: 207-217.

Schoolnik, G.K. (2002a) Microarray analysis of bacterial pathogenicity. *Adv Microb Physiol* **46**: 1-45.

Schoolnik, G.K. (2002b) Functional and comparative genomics of pathogenic bacteria. *Curr Opin Microbiol* **5**: 20-26.

Scott, J.A.G., Hall, A.J., and Dagan, R. (1996) Serogroup-specific epidemiology of *S. pneumoniae*: associations with age, sex and geography in 7000 episodes of invasive disease. *Clin Infect Dis* **1996**: 973-981.

Sebert, M.E., Palmer, L.M., Rosenberg, M., and Weiser, J.N. (2002) Microarray-based identification of *htrA*, a *Streptococcus pneumoniae* gene that is regulated by the CiaRH two-component system and contributes to nasopharyngeal colonization. *Infect Immun* **70**: 4059-4067.

Shakhnovich, E.A., King, S.J., and Weiser, J.N. (2002) Neuraminidase expressed by *Streptococcus pneumoniae* desialylates the lipopolysaccharide of *Neisseria meningitidis* and *Haemophilus influenzae*: a paradigm for interbacterial

competition among pathogens of the human respiratory tract. *Infect Immun* **70**: 7161-7164.

Simberkoff, M.S., Cross, A.P., and Al-Ibrahim, M. (1986) Efficacy of pneumococcal vaccine in high risk patients. *N Engl J Med* **315**: 1318-1327.

Sleeman, K., Knox, K., George, R., Miller, E., Waight, P., Griffiths, D., Efstratiou, A., Broughton, K., Mayon-White, R.T., Moxon, E.R., and Crook, D.W. (2001) Invasive pneumococcal disease in England and Wales: vaccination implications. *J Infect Dis* **183**: 239-246.

Smith, T., Lehman, D., Montgomery, J., Gratten, M., Riley, I.D., and Alpers, M.P. (1993) Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. *Epidemiol Infect* **111**: 27-39.

Snodgrass, J.L., Mohamed, N., Ross, J.M., Sau, S., Lee, C.Y., and Smeltzer, M.S. (1999) Functional analysis of the Staphylococcus aureus collagen adhesin B domain. *Infect Immun* **67**: 3952-3959.

Sorensen, A., Kurland, C.G., and Pedersen, S. (1989) Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* **207**: 365-377.

Spellerberg, B., Rosenow, C., Sha, W., and Tuomanen, E.I. (1996) Pneumococcal cell wall activates NF-kappa B in human monocytes: aspects distinct from endotoxin. *Microb Pathog* **20**: 309-317.

Stern, A., Brown, M., Nickel, P., and Meyer, T.F. (1986) Opacity genes in *Neisseria gonorrhoeae*: control of phase and antigenic variation. *Cell* **47**.

Stibitz, S., Aaronson, W., Monack, D., and Falkow, S. (1989) Phase variation in *Bordetella pertussis* by frameshift mutation in a gene for a novel two-component system. *Nature* **338**: 266-269.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem* **69**: 183-215.

Swiatlo, E., Champlin, F.R., Holman, S.C., Wilson, W.W., and Watt, J.M. (2002) Contribution of choline-binding proteins to celll surface properties of *Streptococcus pneumoniae*. *Infect Immun* **70**: 412-415.

Synder, L.S.A., Butcher, S.A., and Saunders, N.J. (2001) Comparative whole-genome analyes reveal over 100 putative phase-variable genes in the pathogenic *Neisseria spp. Microbiology* **147**: 2321-2332.

Takala, A.K., Vuopio-Varkila, J., Tarkka, E., Leinonen, M., and Musser, J.M. (1996) Sub-typing of common pediatric pneumococcal serotypes from invasive disease and pharyngeal carriage in Finland. *J Infect Dis* **173**: 128-135.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R., and Conway, T. (1999) Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J Bacteriol* **181**: 6425-6440.

Taylor, B.L., and Zhulin, i.B. (1999) PAS domains; internal sensors of oxygen, redox potential and light. *Microbiol Mol Biol Rev* **63**: 479-506.

Teele, D.W., Klein, J.O., and Rosner, B. (1989) Epidemiology of otitis media during the first seven years of life in children in greater Boston: a prospective, cohort study. *J Infect Dis* **160**: 83-94.

Terao, Y., Kawabata, S., Kunitomo, E., Murakami, J., Nakagawa, I., and Hamada, S. (2001) Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the *fba* gene is positively transcribed under the Mga regulator. *Mol Microbiol* **42**: 75-86.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., Durkin, A.S., Gwinn, M.,

Kolonay, J.F., Nelson, W.C., Peterson, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtzapple, E., Khouri, H., Wolf, A.M., Utterback, T.R., Hansen, C.L., McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T., Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K., and Fraser, C.M. (2001) Complete genome sequence of a virulent isolate of Streptococcus pneumoniae. *Science* **293**: 498-506.

Throup, J.P., Koretke, K.K., Bryant, A.P., Ingraham, K.A., Chalker, A.F., Ge, Y., Marra, A., Wallis, N.G., Brown, J.R., Holmes, D.J., Rosenberg, M., and Burnham, M.K. (2000) A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. *Mol Microbiol* **35**: 566-576.

Tomasz, A., Moreillon, P., and Pozzi, G. (1988) Insertional inactivation of the major autolysin gene of Streptococcus pneumoniae. *J Bacteriol* **170**: 5931-5934.

Tong, H.H., Weiser, J.N., James, M.A., and DeMaria, T.F. (2001) Effect of influenza A virus infection on nasopharyngeal colonization and otitis media induced by transparent or opaque phenotype variants of Streptococcus pneumoniae in the chinchilla model. *Infect Immun* **69**: 602-606.

Tu, A.H., Fulgham, R.L., McCrory, M.A., Briles, D.E., and Szalai, A.J. (1999) Pneumococcal surface protein A inhibits complement activation by Streptococcus pneumoniae. *Infect Immun* **67**: 4720-4724.

Tuomanen, E., Tomasz, A., Hengstler, B., and Zak, O. (1985) The relative role of bacterial cell wall and capsule in the induction of inflammation in pneumococcal meningitis. *J Infect Dis* **151**: 535-540.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* **98**: 5116-5121.

Valdivia, R.H., and Falkow, S. (1997) Fluorescent-based isolation of bacteria genes expressed within host cells. *Science* **277**: 2007-2011.

van der Ende, A., Hopman, C.T., Zaat, S., Essink, B.B., Berkhout, B., and Dankert, J. (1995) Variable expression of class 1 outer membrane protein in Neisseria meningitidis is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J Bacteriol* **177**: 2475-2480.

van der Flier, M., Chhun, N., Wizemann, T.M., Min, J., McCarthy, J.B., and Tuomanen, E.I. (1995) Adherence of Streptococcus pneumoniae to immobilized fibronectin. *Infect Immun* **63**: 4317-4322.

Vassalli, J.D., Sappino, A.P., and Belin, D. (1991) The plasminogen activator/plasmin system. *J Clin Invest* **88**: 1067-1072.

Wang, J., Mushegian, A., Lory, S., and Jin, S. (1996) Large-scale isolation of candidate virulence genes of *Pseudomonas aeruginosa* by in vivo selection. *Proc Natl Acad Sci U S A* **93**.

Watson, D.A., Musher, D.M., Jacobson, J.W., and Verhoef, J. (1993) A brief history of the pneumococcus in biomedical research: a panoply of scientific discovery. *Clin Infect Dis* **17**: 913-924.

Ween, O., Gaustad, P., and Manias, J.-M. (1999) Identification of DNA binding sites for ComE, a key regulator of natural competence in *Streptococcus pneumoniae*. *Molecular Microbiology* **33**: 817-827.

Weiser, J.N., Austrian, R., Sreenivasan, P.K., and Masure, H.R. (1994) Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization. *Infect Immun* **62**: 2582-2589.

Weiser, J.N., Markiewicz, Z., Tuomanen, E.I., and Wani, J.H. (1996) Relationship between phase variation in colony morphology, intrastrain variation in cell wall physiology, and nasopharyngeal colonization by Streptococcus pneumoniae. *Infect Immun* **64**: 2240-2245.

Weiser, J.N., and Kapoor, M. (1999) Effect of intrastrain variation in the amount of capsular polysaccharide on genetic transformation of Streptococcus pneumoniae: implications for virulence studies of encapsulated strains. *Infect Immun* **67**: 3690-3692.

Weiser, J.N., Bae, D., Epino, H., Gordon, S.B., Kapoor, M., Zenewicz, L.A., and Shchepetov, M. (2001) Changes in availability of oxygen accentuate differences in capsular polysaccharide expression by phenotypic variants and clinical isolates of Streptococcus pneumoniae. *Infect Immun* **69**: 5430-5439.

Weiser, J.N., Bae, D., Faschin, C., Scamurra, R.W., Ratner, A.J., and Janoff, E.N. (2003) Antibody-enhanced pneumococcal adherence requires IgA1 protease. *Proc Natl Acad Sci U S A* **100**: 4215-4220.

Whitney, C.G., Farley, M.M., Hadler, J., Harrison, L.H., Lexau, C., Reingold, A., Lefkowitz, L., Cieslak, P.R., Cetron, M., Zell, E.R., Jorgensen, J.H., and Schuchat, A. (2000) Increasing prevalence of multidrug-resistant Streptococcus pneumoniae in the United States. *N Engl J Med* **343**: 1917-1924.

Winslow, C.E.A., Broadhurst, J., Buchanan, R.E., Krumwiede, C., Rogers, L.A., and Smith, G.H. (1920) The families and genera of bacteria: final report of the commmittee of the society of American bacteriologists on characterisation and classification of bacterial types. *J Bacteriol* **5**: 191-229.

Wizemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunsch, C., Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E., Gayle, A., Brewah, Y.A., Walsh, W., Barren, P., Lathigra, R., Hanson, M., Langermann, S., Johnson, S., and Koenig, S. (2001) Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection. *Infect Immun* **69**: 1593-1598.

Zhang, J.R., Mostov, K.E., Lamm, M.E., Nanno, M., Shimida, S., Ohwaki, M., and Tuomanen, E. (2000) The polymeric immunoglobulin receptor translocates pneumococci across human nasopharyngeal epithelial cells. *Cell* **102**: 827-837.

APPENDIX A: STATEMENT OF WORK

1. None of the work performed by me for this thesis has previously been submitted for a degree or other qualification.

2. All experiments were performed solely by me except for:
    a. Microarray experiments to identify genes regulated by MgrA, which were performed in collaboration with Dr Elizabeth Joyce, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California, USA.
    b. Ribonuclease protection assays to assess the requirement of RlrA for the expression of pathogenicity islet genes were performed by Dr DL Hava, Department of Molecular Microbiology, Tufts University School of Medicine, Boston, Massachusetts, USA.

3. The work has resulted in two publications, which are included as appendix B in the thesis.
    a. **C Hemsley,** E Joyce, DL Hava, A Kawale and A Camilli. 2003. MgrA, a homologue of Mga, acts as a transcriptional repressor of the *rlrA* pathogenicity islet in *Streptococcus pneumoniae*. J Bacteriol 185:6640-47.
    b. DL Hava, **C Hemsley** and A Camilli. 2003. Transcriptional regulation in the *Streptococcus pneumoniae rlrA* pathogenicity islet by RlrA. J Bacteriol 185:413-21.

APPENDIX
NOT COPIED

ON INSTRUCTION
FROM
THE UNIVERSITY